

STIC-FPAS

From: Goldberg, Jeanine
Sent: Friday, March 12, 2004 10:34 AM
To: STIC-FPAS
Subject: please provide:

CN 1313130 Date: 20010919
Huang, Shi

54 p

THANK YOU

Jeanine Enewold Goldberg
1634
571-272-0743
REM 2E74

U.S. PAT. & TM. OFFICE
MAR 12 2004
MAR 12 2004

FILE COPY

COMPLETED

Copy/Search
Processor: _____
Date Assigned: _____
Date filled: _____
Equivalent found (Y/N): _____
Doc. No.: _____
Country: _____

Translation:
Date logged in: _____
PTO est. words: _____
No. of pages: _____
In-house trans. avail: _____

In-House: _____ Contractor: _____
Translator: _____ Name: _____
Assigned: _____ Priority: _____
Returned: _____ Sent: _____
Returned: _____

FILE 'HOME' ENTERED AT 10:15:12 ON 12 MAR 2004

FILE 'MEDLINE, BIOSIS, CAPLUS, EMBASE, SCISEARCH' ENTERED AT 10:15:26 ON 12 MAR 2004

L1 6786 S MSI OR MICORSATELLITE INSTABILITY
L2 989 S RIZ1 OR RIZ OR RETINOBLASTOMA PROTEIN INTERACTING ZINC
L3 21 S L1 (P) L2
L4 8 DUP REM L3 (13 DUPLICATES REMOVED)

FILE 'STNGUIDE' ENTERED AT 10:17:23 ON 12 MAR 2004

FILE 'STNGUIDE' ENTERED AT 10:22:08 ON 12 MAR 2004

FILE 'MEDLINE, BIOSIS, CAPLUS, EMBASE, SCISEARCH' ENTERED AT 10:22:14 ON 12 MAR 2004

L5 153 S L1 (P) (POLYA?)
L6 68 DUP REM L5 (85 DUPLICATES REMOVED)
L7 1 S L6 AND L2

=>

L Number	Hits	Search Text	DB	Time stamp
1	95	RIZ	USPAT; US-PGPUB; DERWENT	2004/03/12 10:01
2	9	RIZ1	USPAT; US-PGPUB; DERWENT	2004/03/12 10:01
3	95	RIZ or RIZ1	USPAT; US-PGPUB; DERWENT	2004/03/12 10:01
4	3	(RIZ or RIZ1) and "exon 8"	USPAT; US-PGPUB; DERWENT	2004/03/12 10:01
5	11	(RIZ or RIZ1) same deletion	USPAT; US-PGPUB; DERWENT	2004/03/12 10:05
6	13	"retinoblastoma protein" near3 zinc	USPAT; US-PGPUB; DERWENT	2004/03/12 10:06
7	95	("retinoblastoma protein" near3 zinc) or (RIZ or RIZ1)	USPAT; US-PGPUB; DERWENT	2004/03/12 10:06
8	3925	MSI or "microsatellite instability"	USPAT; US-PGPUB; DERWENT	2004/03/12 10:08
9	1	(MSI or "microsatellite instability") same polyA	USPAT; US-PGPUB; DERWENT	2004/03/12 10:08
10	83	(MSI or "microsatellite instability") same delet\$5	USPAT; US-PGPUB; DERWENT	2004/03/12 10:08
11	244	(MSI or "microsatellite instability") same (cancer or tumor)	USPAT; US-PGPUB; DERWENT	2004/03/12 10:09
12	45	((MSI or "microsatellite instability") same (cancer or tumor)) and ((RIZ or RIZ1) or retinoblastoma)	USPAT; US-PGPUB; DERWENT	2004/03/12 10:10
13	42	((MSI or "microsatellite instability") same (cancer or tumor)) and ((RIZ or RIZ1) or retinoblastoma)) and (polymorphism or mutation or deletion or variation)	USPAT; US-PGPUB; DERWENT	2004/03/12 10:12
14	29	((MSI or "microsatellite instability") same (cancer or tumor)) and ((RIZ or RIZ1) or retinoblastoma)) and ((MSI or "microsatellite instability") same (polymorphism or mutation or deletion or variation))	USPAT; US-PGPUB; DERWENT	2004/03/12 10:12

L4 ANSWER 5 OF 8 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 2002:412302 CAPLUS
DOCUMENT NUMBER: 136:363818
TITLE: Application of RIZ gene in detecting and
treating tumor showing MSI positive
INVENTOR(S): Huang, Shi
PATENT ASSIGNEE(S): Peop. Rep. China
SOURCE: Faming Zhuanli Shenqing Gongkai Shuomingshu, 54 pp.
CODEN: CNXXEV
DOCUMENT TYPE: Patent
LANGUAGE: Chinese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
CN 1313130	A	20010919	CN 2001-109921	20010323
PRIORITY APPLN INFO.:			CN 2001-109921	20010323

AB The invention relates to application of adenovirus-mediated RIZ1 gene in treating microsatellite chromosomal DNA instability (MSI)-pos. tumors. The invention also relates to diagnosis of MSI pos. expression in tumor based on the abnormal number of adenosine in poly(A) of RIZ1 gene.

only 1 inventor but
prior support ∴
ref.

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

A61K 48/00

A61P 35/00 C12Q 1/68

[12] 发明专利申请公开说明书

[21] 申请号 01109921.6

[43] 公开日 2001 年 9 月 19 日

[11] 公开号 CN 1313130A

[22] 申请日 2001.3.23 [21] 申请号 01109921.6
[71] 申请人 黄石
地址 100071 北京市丰台区东大街 20 号 29 楼
[72] 发明人 黄石

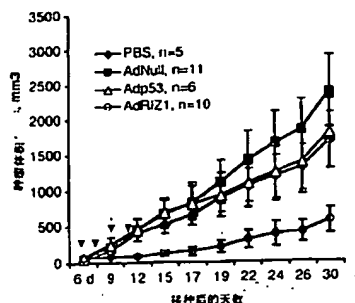
[74] 专利代理机构 北京科龙环宇专利事务所
代理人 孙皓晨

权利要求书 2 页 说明书 44 页 附图页数 7 页

[54] 发明名称 瑞兹(RIZ)基因用于检测和治理卫星染色体不稳定(MSI)阳性肿瘤

[57] 摘要

本发明提供一种抑制卫星染色体不稳定性(MSI)阳性肿瘤生长的方法。这一方法将带有编码有功能的RIZ1多肽的DNA序列转入到MSI阳性肿瘤中,并在肿瘤中表达足够抑制肿瘤生长的RIZ1蛋白质。同时本发明提供一种确定肿瘤MSI状态的方法。这一方法通过确定RIZ1核酸基因中多聚腺嘌呤中腺嘌呤的数目来确定肿瘤是否为MSI阳性。RIZ1基因中存在非正常数目的腺嘌呤表明此肿瘤为MSI阳性。



ISSN 1008-4274

权 利 要 求 书

1、一种用于抑制 MSI 阳性肿瘤的方法。这种方法由编码 RIZ1 多肽的核酸分子组，能介导到肿瘤中，并在肿瘤中表达足够量的这种多肽，以达到抑制肿瘤的效果。

2、权利要求 1 所要求的方法，所述肿瘤在 RIZ 多聚 A 位点有非正常数目的腺嘌呤。

3、权利要求 2 所要求的方法，其中 RIZ 多聚 A 位点在 RIZ (A) 8 序列中。

4、权利要求 2 所要求的方法，其中 RIZ 多聚 A 位点在 RIZ (A) 9 序列中。

5、权利要求 2 所要求的方法，其中非正常数目的腺嘌呤包括数目增加。

6、权利要求 2 所要求的方法，其中非正常数目的腺嘌呤包括数目减少。

7、权利要求 1 所要求的方法，其中所述核酸分子组在肿瘤中表达 RIZ1 导致肿瘤细胞凋亡。

8、权利要求 1 所要求的方法，其中 RIZ1 多肽含有与序列号 99 中有 95% 以上相同的 PR 氨基酸序列。

9、权利要求 8 所要求的方法，其中 RIZ1 多肽包括序列号 99 顺序。

10、权利要求 8 所要求的方法，其中 RIZ1 多肽含有与序列号 4 有 80% 以上相同的氨基酸序列。

11、权利要求 10 所要求的方法，其中 RIZ1 多肽包括序列号 4 顺序。

12、权利要求 1 所要求的方法，其中核酸序列处于一个病毒载体中。

13、权利要求 12 所要求的方法，其中病毒载体是一个腺病毒载体。

14、权利要求 1 所要求的方法，其中核酸以肿瘤内或肿瘤外周注射的方式导入肿瘤细胞。

15、权利要求 1 所要求的方法，其中 MSI 阳性肿瘤由部分结肠癌，胃癌，或子宫内膜癌组成。

16、权利要求 15 所要求的方法，其中 MSI 阳性肿瘤是家族型非息肉性结肠肿瘤。

17、一种测定 MSI 阳性状态的方法，该方法通过测量肿瘤中 RIZ 基因上多聚 A 位点上腺嘌呤的数目来决定 MSI 状态，RIZ 基因上中非正常数目的多聚 A 表明此肿瘤为 MSI 阳性。

18、权利要求 17 所要求的方法，其中 RIZ 多聚 A 位点在 RIZ (A) 8 序列中。

19、权利要求 17 所要求的方法，其中 RIZ 多聚 A 位点在 RIZ (A) 9 序列中。

20、权利要求 17 所要求的方法，其中非正常数目的腺嘌呤包括数目增加。

21、权利要求 17 所要求的方法，其中非正常数目的腺嘌呤包括数目减少。

22、权利要求 17 所要求的方法，其中 MSI 阳性肿瘤由部分结肠癌，胃癌，或子宫内膜癌组成。

23、权利要求 22 所要求的方法，其中 MSI 阳性肿瘤是家族型非息肉性结肠肿瘤。

说明书

瑞兹 (RIZ) 基因用于检测和治疗卫星染色体不稳定 (MSI) 阳性肿瘤

一、发明领域

本专利发明涉及肿瘤领域, 具体是利用 RIZ 肿瘤抑制基因来检测和治疗卫星染色体不稳定阳性 (MSI+) 肿瘤。

二、背景材料

成视网膜细胞瘤蛋白 (Retinoblastoma 或 RB) 相关锌手指 (Zinc Finger) 基因瑞兹 (RIZ) 是一个潜在的肿瘤抑制基因, 它属于 PR (PRDI-BFI-RIZ-1) 或 SET (Suvar3-9, Enhancer-of zeste, Trithorax) 相关蛋白家族, 在染色体介导的基因激活和失活过程中起作用。此家族与染色体相关的基因激活或失活有关系。PR/SET 区域家族在人类肿瘤中起重要作用, 因为此家族中多个成员在肿瘤中有遗传性改变。RIZ 基因中的 PR 区域是与其它蛋白质相作用的位点并能与其自身羧基 (C-端) 的一个区域相结合。

RIZ 基因一般有两个蛋白质产物, RIZ1 和 RIZ2, 其区别在于氨基端是否有 PR 区域。RIZ1 (带有 PR 区域) 产物被认为是一个位于 1 号染色体短臂 36 区 (1p36) 的很强的潜在肿瘤抑制基因。1p36 区在人类十二种以上肿瘤中有丢失。在多种肿瘤和肿瘤细胞株中, RIZ1 而非 RIZ2 基因不再表达。这些肿瘤包括乳腺癌, 肝癌, 结肠癌, 和神经内分泌系统肿瘤 (He 等, Cancer Research. 58:4238-4244, 1998; Jiang 等, Int. J. Cancer. 83:541-544, 1999)。这些肿瘤的特性是 RIZ1 基因失活, 而非 RIZ1 基因突变引起蛋白质结构变化。在这些肿瘤细胞中, 外源性 RIZ1 基因表达导致细胞分裂周期在 G2/M 期阻断, 细胞凋亡, 或二者兼有。然而在体内 RIZ1 表达在各种肿瘤中的作用还未被证实。

肿瘤发生是由于正常细胞内致癌基因遗传性突变而形成的这一观点已经被公认。这些突变引起遗传不稳定性。在肿瘤中两种主要遗传不稳定性已经被确认, 染色体不稳定性 (CIN) 和卫星染色体不稳定性 (MSI)。CIN 的主要特征是染色体异倍体和丧失异源性 (LOH), 相反 MSI+ 肿瘤一般是染色体双倍体, 而在许多简单的重复序列, 卫星染色体中出现众多不稳定性。

卫星染色体不稳定性被认为是由于细胞内脱氧核糖核酸 (DNA) 碱基错误配对修复系统失灵而造成的。这个系统在正常情况下, 识别并修复在 DNA 复制中错误使用的碱基或碱基错位。碱基配对错误修复机制的丧失引发突变进化和选择过程, 形成肿瘤。MSI 阳性的肿瘤形成机制与卫星染色体重复序列错位突变有关。当这些重复序列位于目标基因的编码序列中时, 错位突变会引发肿瘤形成。

Adnull 或对照缓冲液，两天后进行 RIZ1, p53, 或 Apoptag 细胞凋亡原位检测 (TUNEL 分析)。

发明具体描述

本发明提供一种抑制 MSI 阳性肿瘤生长的方法，包括在肿瘤中导入编码有功能的 RIZ 多肽的 DNA 序列，并在肿瘤中表达足够量的 RIZ 多肽以抑制肿瘤生长。象在发明实施例中展示的那样，RIZ1 多肽在体内和体外抑制 MSI 阳性肿瘤细胞和肿瘤生长。在具有 RIZ 多聚腺嘌呤（多聚 A）错位突变的 MSI 阳性肿瘤中，RIZ1 多肽在体内和体外还具有诱导细胞凋亡的功能。

这里所指的卫星染色体不稳定性阳性肿瘤细胞或 MSI 阳性肿瘤细胞指细胞中至少 20% 的卫星染色体标记顺序与正常细胞相比在长度上出现异常，更好的 MSI 阳性肿瘤是有 40% 以上测试过的卫星染色体出现异常；最好有 60% 以上出现异常。这些数据至少需在测定 5 个以上标记顺序后得出。理想的情况是这些细胞在 RIZ 基因中有卫星染色体出现异常。然而由于在 MSI 阳性肿瘤中，RIZ1 多聚 A 错位突变出现的比例很高，不必要事先测定 RIZ1 是否有突变就可预测在大多数人群中会有有效的治疗作用。

卫星染色体是简单的遍布于染色体中的重复序列，包括一个，两个，三个，或四个核苷酸。已经知道的有一系列标记顺序是 MSI 阳性的标志（见 Boland 等，Cancer Res. 58:5248-5257, 1998）。而且市场上有现成的分析卫星染色体的引物。这些标记包括 Boland 等在上述文献中提到的标志物组，单核苷酸重复的 BAT25（基因库 Genbank 序列 9834508）和 BAT26（基因库 Genbank 序列 9834505），双核苷酸重复的 D5S346（基因库 Genbank 序列 181171），D2S123（基因库 Genbank 序列 187953），和 D17S250（基因库 Genbank 序列 177030）。其它合适的标记序列亦有描述，如 Boland 等在上述文献中提到的，以及 Mao 等在 Proc. Natl. Acad. Sci. USA 91:9871-9875, 1994 中提到的序列。在本发明中使用的标记序列是上述提到的五个标志物组。

卫星染色体不稳定性与很大一部分自发性和遗传性肿瘤有关。本发明方法用于检测 MSI 状态和抑制自发性和遗传性两种肿瘤。

具体的讲，MSI 阳性已在头颈癌（如未或极少分化的头颈癌），结肠癌，甲状腺癌，食道癌，胃癌，前列腺癌，子宫内膜癌，卵巢癌，子宫颈癌，乳腺癌，皮癌（扩散性皮癌），小细胞肺癌，非小细胞肺癌，和慢性骨髓癌（成骨细胞危机）等（见 Claij 等综述，Exp. Cell. Res. 246:1-10, 1999）。

多数家族性的非息肉性结肠癌（HNPCC）也存在卫星染色体不稳定性。HNPCC 是一种家族性肿瘤，占总结肠癌的 10%。HNPCC 的主要特点是肿瘤发生早，并同时有其它部位原发瘤，包括结肠，子宫内膜，卵巢，小肠，胃，尿道，和皮肤等。本发明方法尤其对 HNPCC 的检测和治疗有特别效用。

除了避免发生突变外, DNA 错配修复还决定用于肿瘤化疗的 DNA 致变试剂的毒性。由于多种化疗药品, 甲基化的试剂, 如 N-methyl-N'-nitro-nitroguanidine, N-methyl-N-nitrosourea, Streptozocin, temozolonide, 和 decarbazine; 铂化试剂, 如 Cisplatin 和 Carboplatin; 碱基类似物, 如 6-thioguanine; 及其他化疗试剂, 如 busulfan, etoposide, 和 doxorubin 等的杀死肿瘤细胞功能需要有正常的错配修复系统, 因此在 MSI 阳性的肿瘤细胞中进行化疗效果不明显。

利用肿瘤抑制基因进行肿瘤基因治疗是一种简单和能取代化疗和放疗的低毒性产品。正在进行的几个临床试验将肿瘤抑制基因 P53 或 RB 基因利用逆转录病毒或腺病毒载体导入到带有上述基因突变的肿瘤中(见 Roth 等, Oncology, 135:148-153, 1999)。然而在 MSI 阳性肿瘤中基本没有 RB 或 P53 基因缺陷, 因此不能用这些基因进行基因治疗。至今还没有发现一种有效用于 MSI 阳性肿瘤的基因治疗方法。

因此有必要开发治疗 MSI 阳性肿瘤的基因治疗方法, 并发现一种可靠的检测 MSI 阳性肿瘤的标记, 这个专利发明就能在这方面填补空白并有更多建树。

发明综述

本发明提供一种抑制卫星染色体不稳定性(MSI)阳性肿瘤生长的方法。这一方法将带有编码有功能的 RIZ1 多肽的 DNA 序列转入到 MSI 阳性肿瘤中, 并在肿瘤中表达足够抑制肿瘤生长的 RIZ1 蛋白质。

同时此发明提供一种确定肿瘤 MSI 状态的方法。这一方法通过确定 RIZ1 核酸基因中多聚腺嘌呤中腺嘌呤的数目来确定肿瘤是否为 MSI 阳性。RIZ1 基因中存在非正常数目的腺嘌呤表明此肿瘤为 MSI 阳性。

图例注释

图 1 表示人的 RIZ 蛋白质氨基酸顺序(标识为 hRIZ; SEQ ID No. 4)与大鼠 RIZ 蛋白质氨基酸顺序(标识为 rRIZ; SEQ ID No. 2)相关性比较。共同序列已标明。使用一氨基酸代号。在人与大鼠中不相同的氨基酸标识为“.”

图 2 在裸鼠异体肿瘤模型中利用 AdRIZ1 腺病毒载体对卫星染色体不稳定性阳性(MSI+)的结肠瘤进行基因治疗。将带有 RIZ1 基因突变的人类结肠瘤 HCT116 细胞注射到裸鼠的侧腹部, 治疗前, 带有肿瘤的裸鼠根据肿瘤大小进行随机分组。治疗中在肿瘤内或肿瘤外周分别注射缓冲液(PBS), 对照空白腺病毒载体(Adnull), 带 p53 基因的腺病毒载体(Adp53), 或带 RIZ 基因的腺病毒载体(AdRIZ1)。共注射四次, 每次剂量为每个动物 8×10^{10} 病毒颗粒。箭头指示注射时间, 每周两次测量肿瘤体积, 连续测量四周。

图 3 在腺病毒转导的肿瘤细胞中进行 RIZ1, p53, 和细胞凋亡的免疫组织化学染色分析。在体内已形成肿瘤的 HCT116 细胞在注射 AdRIZ1, Adp53,

这里所述的肿瘤是指在局部肿瘤细胞的生长，可以是肿瘤原发位点或转移位点。肿瘤细胞是指恶性转化细胞，或在肿瘤，或转移位点，或从肿瘤或转移位点取出的肿瘤细胞。从肿瘤或转移位点取出的肿瘤细胞可在体外培育多代。

这里所述的抑制肿瘤生长是指减缓肿瘤细胞生长速度，阻断肿瘤细胞生长或在肿瘤内杀死肿瘤细胞，因此而使肿瘤生长速度与未处理的对照相比有所减缓。抑制生长亦指肿瘤的减小或消失或降低转移能力。本技术领域的专业人士可用各种指标来确定肿瘤生长是否被抑制。

比如抑制肿瘤生长可以阻断细胞生长周期。如实施例 1 所示，外源的 RIZ1 表达能使 MSI 阳性肿瘤细胞生长周期停留在 G2/M 期，不管这些细胞是否带有 RIZ 基因多聚腺嘌呤重复序列错位突变。

抑制肿瘤生长可以以直接或间接方式测定肿瘤大小来确定。如实施例 2 所示，外源的 RIZ1 表达能使 MSI 阳性肿瘤缩小。在肿瘤病人中，此类测量可用各种已知的显影方法如核磁共振显影，计算机轴向全息显影，或 X 光照影。肿瘤生长也可间接测量。如测定血液中胚胎原抗原（CEA），前列腺特异抗原（PSA），或其它与肿瘤生长相关的肿瘤特异抗原。抑制肿瘤生长一般会导致延长生存和/或有益病人的健康，提高生活质量。

如下实施例 1 和实施例 3 所示，在 MSI 阳性肿瘤细胞株或肿瘤中，特别是在有 RIZ 基因多聚腺嘌呤重复序列错位突变的 MSI 阳性肿瘤中，表达外源的 RIZ1 会引发细胞凋亡。这里用的引发细胞凋亡是指刺激以 DNA 断裂为特征的一种程序性细胞死亡。细胞凋亡可以以已知的方式来确定。如实施例 3 所示，现有各种测试试剂盒利用原位组织化学杂交的方式来确定 DNA 片段的存在（如 Apoptag 试剂，由 Intergen 提供）。例外，如实施例 1 所示细胞凋亡亦可用荧光细胞分检（FACS）的方法来确定。在 FACS 分析中存在的低于 G1 期 DNA 含量的细胞群表示这些处于细胞凋亡的细胞有 DNA 断裂发生。

本发明适用于在 MSI 阳性肿瘤中引入能编码 RIZ1 多肽的核酸分子，并表达有效剂量 RIZ1 蛋白质来治疗肿瘤。这里所用的 RIZ1 多肽指具有 RIZ1 氨基酸序列的多肽。RIZ1 多肽亦指这里所列 4 号序列（SEQ ID No. 4）或其功能部分相同，或与其有一个或很小差异的序列，只要此序列具有抑制 MSI 阳性肿瘤或肿瘤细胞生长的功能。很小差异包括在 4 号序列中进行一个或多个氨基酸序列的增加，减少，或替代，但并不特别影响其抑制肿瘤或肿瘤细胞生长的功能。

克隆的人类 RIZ 基因序列（所列 3 号序列）编码 4 号序列中的氨基酸序列。克隆的大鼠 RIZ 基因序列（所列 1 号序列）编码 2 号序列中的氨基酸序列；小鼠 RIZ 基因序列（104 号序列）已在美国专利号 6069231，5831108，和 5811304 以及在 Buyse 等 Proc. Natl. Acad. Sci. USA 92:4467-4471，1995 中描述过。

一种可称作 RIZ1 的多肽可与 4 号序列中描述的氨基酸序列有 70% 的相同之处，最好具有 75% 以上，包括 80%，85%，90%，95%，98%，99%，或更多相同之处。理想的例子是从另一种哺乳动物中克隆的 RIZ 基因。如图一所示，人的 RIZ 基因（所列 4 号序列）与大鼠 RIZ 基因序列（所列 2 号序列）有 84% 的相同之处。

已在本领域广为接受的观点是进化中保留下来的氨基酸或氨基酸区域比其它氨基酸或氨基酸区域更可能有更重要的生物学功能。因此可以推测用一个不具保留系数的氨基酸来替换一个在所有种类的 RIZ 多肽中保留下来的氨基酸将会带来功能性灾难，而用一个不具保留系数的氨基酸替换一个在各种类的 RIZ 多肽中变化大的氨基酸将不会对其功能有重大影响。

替换 2 号序列中的氨基酸可以是保留性的，也可以是非保留性的。保留性替换包括，但不限于用一个非极性氨基酸替代另一个非极性氨基酸，如用异亮氨酸，缬氨酸，丙氨酸，脯氨酸，苯丙氨酸，色氨酸，或蛋氨酸来替代亮氨酸；用一个带电的氨基酸替代类似带电的氨基酸，如天门冬氨酸替代谷氨酸，赖氨酸或组氨酸替代精氨酸；用一个不带电荷的非极性氨基酸替代另一个不带电荷的非极性氨基酸（如用甘氨酸，苏氨酸，酪氨酸，胱氨酸，天门冬酰胺或谷酰胺替代丝氨酸；或用不同组分但带有相似大小和形状的氨基酸替代另一个，如用丙氨酸替代丝氨酸，用精氨酸替代蛋氨酸，或苯丙氨酸替代酪氨酸。

在 RIZ 多肽中增加序列包括，但不限于在氨基端，羧基端，或中间增加标记序列，此标记序列包括，但不限于功能序列，多聚组氨酸序列，GST 序列，或类似序列及可分离的序列。这些序列可用来促进重组 RIZ1 表达或鉴别，包括分离。

对序列 4 中的 RIZ1 多肽修饰的事例包括正常的序列 4 的染色体正常异体，如序列 102 和 103 中所示 RIZ1 头三个氨基酸变化。另一个染色体正常异体的事例为 T969 变成 A969（969 位的胸腺嘧啶变成腺嘌呤），导致氨基酸序列 283 由 D 变成 E。T969 异体估计比 A969 异体发生的几率高一倍。序列 4 中的另一个染色体正常异体是 704 位的脯氨酸丢失。

在一种应用中，用在本发明使用方法在肿瘤中表达的 RIZ1 多肽由序列 4 的 PR 区域组成（此序列亦称为序列 99），或由与此序列具有 95% 以上相同的序列，如大鼠的 RIZ1 PR 区域组成。

在 RIZ1 氨基端有一个 100 多个氨基酸的区域叫 PR 区域，因为它与 PRD1-BF1 和 Blimp-1 蛋白的 N-端具有 42% 的同源性。PR 区域亦与 Evi-1 蛋白的 N-端（见 Morishita 等，cell 54:831-840, 1998; Morishita 等，Oncogene, 5: 936-971, 1990）和 C. Elegans（细小线虫）的 egl-43 蛋白的 N-端（见 Garriga 等，Genes Devel. 7:2097-2109, 1993 中图 13）具有同源性。

PR 区域已被证明为 RIZ1 负调控作用所必需（见 He 等, *Cancer Res.* 58: 4238-4244, 1998），并可能参与染色体介导的转录激活或抑制（Xie 等, *J. Biol. Chem.* 272:26360-26366, 1997; Huang 等, *J. Biol. Chem.* 273:15933-15939, 1998）。

总体来讲, PR 区域涵盖 100-120 个氨基酸, 有三个高度保留的位点, A, B, C 群。每个有 10-12 个氨基酸, 中间被 20-35 个氨基酸分开。每个 A, B, C 群由单一外显子编码。大鼠的 RIZ1 PR 区域序列 2 中(氨基酸 36-151)和人类 RIZ PR 区域(序列 4 中 37-152, 亦同序列 99)基本相同, 只是人的在 70 位氨基酸上是赖氨酸, 而在相同位置上(69 位), 大鼠的是精氨酸。另外从 DNA 分析中得出编码小鼠 PR 区域 B 和 C 群(序列号 104 和 105)序列中推导的氨基酸顺序(即序列 105 中的前 75 个氨基酸)与人类 RIZ1 中 PR 区域(即人类 RIZ1 序列 99 中 42-116)完全相同。

RIZ1 多肽还有几个有趣的地方, 包括与 E1A 肿瘤蛋白相似的 RB 结合位点; 8 个锌手指序列; 羧基端 PR 结合位点。N-端 PR 区域与 C-端 PR 结合区域的相互作用可能为 RIZ1 形成单体或异体寡聚物所必需, 亦可能参与与其它蛋白间的相互作用。

通过此发明所提供的方法, 其它本行同仁会很容易决定一个经修饰序列 4 或其一部分后得到的 RIZ1 多肽是否具有抑制 MIS 阳性肿瘤或肿瘤细胞的功能。如实施例 1 和实施例 3 所示, 一个能抑制体外 MSI 阳性肿瘤细胞生长的 RIZ1 多肽亦具有抑制 MIS 阳性肿瘤生长的功能。因此, 可将一个修饰序列 4 或其一部分后得到的可编码 RIZ 多肽的核酸序列导入 MSI 阳性肿瘤或 MSI 阳性肿瘤细胞, 如 HCT116, 来决定其是否具有抑制肿瘤生长的功能。如果它仍具有抑制功能, 则此核酸序列适用于本发明提供的方法。

可将带有编码 RIZ 多肽的核酸序列装入适当的载体中用于在哺乳类动物中表达。合适的载体包括, 但不限于病毒载体, 如逆转录病毒(复制缺陷的 MuLV, HTLV, 和 HIV 载体), 腺病毒载体, 腺相关病毒载体, 疱疹病毒载体, 和非病毒载体, 如病毒基因组, 质粒 DNA, 和噬菌体质粒等(参见 Kaplitt 和 Loewy, *Viral Vectors. Gene Therapy and Neuroscience Applications.* Acad. Press, San Diego, CA 1995; Chang, *Somatic Gene Therapy*, CRC Press, Boca Raton, FL, 1995)。克隆这些目的核酸的方法在同仁中已广为所知。

腺病毒载体具有特殊的优越性, 因为它们可以转导处于生长期和静止期的细胞, 并能在体外大量培养, 而且腺病毒载体不整合到细胞宿主的基因组上, 具有更好的安全性。腺病毒载体已经被成功应用于转导 P53 肿瘤抑制基因到肿瘤中, 具有低毒性, 高水平基因表达, 和明显治疗效果(见 Swisher 等, *J. Natl. Cancer. Inst.* 91:763-771, 1999; Clayman, *Seminars in Oncology* 2788:39-43, 2000)。

合适的话，一个病毒载体或其它载体可通过在载体中加入组织特异性启动子或增强子的方法用来在特定的组织中表达编码 RIZ 基因的核酸（如 Dai 等, *Proc. Natl. Acad. Sci.* 89:10892-10895, 1992）。或者编码 RIZ 基因的核酸不用载体而通过合适的可帮助转录和翻译的核酸序列达到有效表达。

本技术领域中的技术人员可决定合适的可传递核酸分子的配方，如在载体类别（侵染性和非侵染性）和传递方式上作出合适选择。合适的配方包括，但不限于如将核酸包含于脂质体中，将核酸与高价阳离子聚合物混合，将核酸分子与靶向性分子组合（如抗体，配体，融合性多肽，或 HIV Tat 蛋白质等）。基因治疗方法，包括选择合适载体，启动子，和配方可在各种综述中找到，如 Anderson, *Nature* 392:25-30, 1998。

编码 RIZ 多肽的核酸可以各种途径转入到病人体内从而在肿瘤中表达有效剂量的蛋白质。在最佳状态下，核酸转入的途径是局部性的，如通过肿瘤内或外周注射的方式达到。根据肿瘤的位置，局部导入可与手术同步进行或通过显影方式指导导入仪器到肿瘤位点（如 Swisher 等, *Supra*, 1999）。局部导入的优势是载体没有稀释，因此会增加大多数肿瘤细胞能接触到此核酸的机会。

导入编码 RIZ 多肽的核酸的另一种方式是系统转导，如腹下，静脉注射，或皮下注射。系统转导方法对治疗扩散的肿瘤有优势。

受体介导的 DNA 导入方式在将编码 RIZ 多肽的核酸转入到非肿瘤位点时有优势。如一个病毒载体可与组织特异或肿瘤特异性配体或抗体以桥状连接分子结合，导入体内后，病毒颗粒将在体内循环直到找到那些带有特异性靶体的细胞进行结合和基因转导。

本发明亦提供一种检测 MSI 阳性特征的方法。其方法是确定肿瘤细胞中 RIZ 核酸中多聚腺嘌呤中腺嘌呤的个数。非正常数目的腺嘌呤数表明此肿瘤为 MSI 阳性。

如本文所述，在极大多数 MSI 阳性肿瘤细胞和细胞株中，包括结肠癌，胃癌，和子宫内膜癌细胞株中，存在多聚腺嘌呤错位突变。相反在 MSI 阴性肿瘤细胞中，没有检测到 RIZ 多聚腺嘌呤错位突变。因此检测到肿瘤中含有 RIZ 基因多聚腺嘌呤错位突变极高地预测此肿瘤为 MSI 阳性肿瘤。

通过检查 RIZ1 cDNA 顺序发现在外显子 8 区域有两处容易产生多聚腺嘌呤（PolyA）变化的高突变区。一个是在序列 3 的 4393-4400 中的 A8，另一个是在序列 3 的 4582-4590 中的 A9。RIZ 基因 A8 多聚腺嘌呤序列位于含有锌手指最多位点的 C-端的上游，此处错位突变造成翻译时锌手指断开。A9 序列在此同一锌手指的下游，此处错位突变将丧失 RIZ 蛋白 C-端 PR 区域结合位点。

确定 RIZ 基因中 A8 或 A9 序列中腺嘌呤多少的方法在同行的同仁中广泛知晓。例如，PCR 扩增 RIZ 染色体 DNA，RIZ 转录 DNA，然后进行单一核苷酸多

型性 (SSCP) 分析, 或直接进行 DNA 测序。在此实施例 1 和实施例 2 中有具体检测多聚 A 数目的方法。

检测的 DNA 样品最好直接来源于肿瘤, 如活检组织。新鲜或处理后的组织切片亦可用来检测。在特定的应用方法中, 可分析尿液, 血液, 淋巴中, 或培养后的肿瘤细胞。

由前所述, MSI 阳性肿瘤比 MSI 阴性肿瘤对某些化疗不敏感。因此治疗前确定肿瘤是否为 MSI 阳性很重要, 这样有利于选择最佳治疗方案。象这里描述的, MSI 阳性肿瘤能被外源表达的 RIZ1 基因所抑制。因此 RIZ 基因治疗可作为化疗外的其它选择, 或与其它治疗方法一起使用。

下述实施例只为了说明但不限制此发明的应用。

实施例 1、在 MSI 阳性肿瘤中 RIZ1 基因突变和外源 RIZ 基因表达的作用

这个实施例表明在多种细胞起源的 MSI 阳性肿瘤中存在有高水平的 RIZ1 基因多聚 A 序列错位突变。此实施例亦表明在含有 RIZ 基因多聚 A 序列错位突变的 MSI 阳性肿瘤中导入能表达编码 RIZ 多肽功能的核酸能够导致细胞周期阻断和肿瘤细胞凋亡。

实验材料和方法

组织样品和细胞株 共研究了 22 个 MSI 阴性肿瘤细胞, 其中 8 个的选择是因为它们曾被确定有染色体不稳定性 (CIN) 表型, 它们包括在 1 号染色体两个相邻的标记点 D1S228 (32.4 摩尔根 (cm)) 和 D1S507 (36.2 cm) 上丧失异源性 (LOH) (Canzian 等, Cancer Res. 56:3331-3337, 1996)。另 14 个肿瘤细胞在 1 号染色体上的 LOH 情况未知。3 个 MSI 阴性肿瘤细胞株由美国标准细胞组织中心 (ATCC) 获得, 这三个株是 MDAMB231, MDAMD435S, 和 SKBR3 (乳腺癌), 以前曾检查过这些细胞中 RIZ 的表达 (He 等, Cancer Res. 58:4238-4244, 1998)。

另外研究了 24 个从家族型非肉瘤性结肠癌 (HNPCC) 病人身上取得的 MSI 阳性肿瘤 (Aaltoren 等, N. Eng. J. Med. 338:1481-1487, 1998), 还研究了 11 个由 ATCC 获得的 MSI 阳性肿瘤细胞株。它们是 DLD1, LS411N, SW48, HCT116, LoVo, HCT15, 和 LS174T (结肠癌); MDAH2774, 和 SK-OV3 (卵巢癌); AN3CA (子宫内膜癌); 和 DU145 (前列腺癌)。

丧失异源性 (LOH) 分析 利用荧光标记卫星染色体方法分析正常和原发肿瘤细胞对。引物序列由基因组数据库获得 (<http://www.gdb.org>)。卫星染色体扩增条件为: 在 15 μ l 反应液中有 10 ng 相应的染色体 DNA, 8 pMol 引物 (5' 引物有荧光标记), 100 nM 的四种脱氧核苷酸三磷酸 (dNTPs), 0.6 单位 (U) DNA 聚合酶 (AmpliTaq Gold, PE Biosystems, Foster City, CA), 10 mM Tris.HCL (pH 8.3), 50 mM 氯化钾 (KCl), 2 mM 氯化镁 ($MgCl_2$)。PCR 产物在 377XL 序列分析仪上进行分析 (PE Biosystems)。染色体各体的大

小和荧光强度由 GENESCAN 和 GENOTYPER 软件 (PE Biosystems) 自动检测。由分析染色体各体的强度比例关系即可计算 LOH 的程度 (Canzian 等, Cancer Res. 56:3331-3337, 1996)。如果一个样品染色体各体的比例小于 0.67 或大于 1.5, 则被认为存在 LOH。

由于很多 HNPCC 肿瘤是 MSI 阳性, 单一核苷酸多样性 (SNPs) 方法亦可用来分析 24 个 HNPCC 正常和肿瘤 DNA 对。引物序列由人类 SNP 数据库获得 (<http://www.genome.wi.mit.edu/snp/human/index.html>)。SNP 扩增条件为: 在 25 μ l 反应液中有 20 ng 相应的染色体 DNA, 8 nMol 相应的引物, 100 nM 的四种脱氧核苷酸三磷酸 (dNTPs), 1.0 单位 (U) DNA 聚合酶 (AmpliTaQ Gold, PE Biosystems, Foster City, CA), 10 mM Tris.HCL (pH 8.3), 50 mM 氯化钾 (KCl), 2 mM 氯化镁 ($MgCl_2$)。PCR 产物经核酸外切酶 1 和虾碱性磷酸化酶处理后, 直接用其中一个引物进行单方向 DNA 测序。DNA 测序不好的样品再用另一个引物进行测序。LOH 确定方式与上述卫星染色体分析方法相似。

突变分析 目标基因的突变经直接测染色体 PCR 产物顺序进行。为方便直接测序, 所有引物都带有 M13 前向 (TGTAACGACGGCCAGT, 序列号 107) 和 M13 反向 (CAGGAAACAGCTATGACC, 序列号 108) 序列。PCR 条件是: 在 25 μ l 反应液中有 20 ng 相应的染色体 DNA, 8 nMol 相应的引物, 100 nM 的四种脱氧核苷酸三磷酸 (dNTPs), 1.0 单位 (U) DNA 聚合酶 (AmpliTaQ Gold, PE Biosystems, Foster City, CA), 10 mM Tris.HCL (pH 8.3), 50 mM 氯化钾 (KCl), 2 mM 氯化镁 ($MgCl_2$)。PCR 产物经核酸外切酶 1 和虾碱性磷酸化酶处理后, 直接用其中一个引物进行单方向 DNA 测序。

基因表达分析 用于表达的 RNA 先用 RNAase 试剂盒 (Qiagen, Chatsworth, CA) 抽提。反转录时用 Superscript RT (Life Technologies, Rockville, MD), 然后用 GeneAmp Gold RNA PCR 试剂盒 (PE Biosystems) 进行 cDNA 扩增。用于确定 RIZ1 和 RIZ2 的特异性引物已有描述 (见 He 等, Supra, 1998)。对照为 3-磷酸甘油醛脱氢酶 (GADPH) 特异性引物 (Leygue 等, Cancer Res. 58:1348-1352, 1998)。

免疫沉淀和免疫转移膜杂交分析 根据 Buyse 等在 Proc. Natl. Acad. Sci. 92:4467-4471 (1995) 中描述的方法进行免疫沉淀和免疫转移膜杂交分析。用对应于 RIZ2 N-端的 RIZ 抗体血清 1637 或免疫前血清 (对照) 将细胞裂解物进行免疫沉淀。沉淀物经 5% 的 SDS 电泳中分离, 然后用对应于 RIZ2 N-端的 RIZ 抗体血清 KGSE 进行免疫转移膜杂交分析。

分析外源性 RIZ1 在细胞中的表达对细胞周期和细胞凋亡的影响 外源性 RIZ1 基因的表达对细胞周期和细胞凋亡的作用分析方法详见 He 等, Supra, 1998。具体方法简述如下。先将 2×10^5 结肠癌细胞种在 6 cm 的培养皿中, 然后用重组

腺病毒载体侵染（病毒 / 细胞比率为 100）。48 小时后，对细胞进行 DNA 量分析。

实验结果

1 号染色体短臂丢失图谱 很多自发性结肠癌在 1 号染色体端点上有大致 40cm 大小的丢失。在这个区域有些肿瘤又回到具有异源性（见 Chadwick 等, Proc. Natl. Acad. Sci. 97:2662-2667, 2000 中图 1）。在这个区域有几个潜在的肿瘤基因和肿瘤抑制基因，包括在 46.2 cm 的 PAX7, 在 46.2 cm 的 PLA2, 在 52.4 cm 的 E2F2, 和在 71 cm 的 MYCL。与 MSI 阴性肿瘤相比，有共同丢失区域的肿瘤更限于 MSI 阳性的 HNPCC 病人肿瘤。虽然大多数位于 1 号染色体末端的 SNP 标志物确定性不强，但位于 32.2 cm 的标志物 WIAF-481 是一个常见的丢失位点，它在 53%（15 个中有 8 个，8/15）已确定的 HNPCC 肿瘤中有 LOH。染色体各体在 SNP 中的不平衡可以重复检测到，表明这个位点在检测染色体丢失图谱，特别是 MSI 阳性肿瘤中很有价值。相比之下，在 22.9 cm 处的 D1S450（1/11）和在 32.4 cm 处的 D1S228（1/8）LOH 的比率就低得多。

确定 RIZ 是 1 号染色体短臂变化的潜在基因 我们对位于 1pTel 32.2 cm 处的功能基因进行突变分析，RIZ 基因位于基因图 99 号的 32.2 cm 处（<http://www.ncbi.nlm.nih.gov/genemap99>）。由酵母人工染色体分析得出 RIZ 离 D1S228 标记点 370 kb 远（Leygue 等, Supra, 1998）。RIZ 基因以 3' 到 5' 方向远离 1 号染色体短臂末端。在 HNPCC 肿瘤中，从 WIAF-481 到 RIZ 3' 到 RIZ 5' 方向 LOH 频率逐渐减少。对 RIZ 基因内三个位点，包括外显子 8 中密码子脯氨酸 704 丢失，在 4 号内显子中的 CA 重复序列，和在 4 号外显子边的 SNP（在 4 号外显子 3' 端 18 个碱基的 GAT 变成 GAC）进行了多型体分析。这些 RIZ 基因标记的 LOH 频率分别为 21%（3/14），9%（1/11），和 0%（0/8）。5393T 肿瘤表现在 8 号外显子 Pro-704 有 LOH，但在其它位点没有变化，与其在 RIZ 基因中丢失断开位点相一致。在 MSI 阴性但未事先做 1 号染色体短臂 LOH 分析选择的肿瘤中发现这些位点的 LOH 频率分别为 30%（4/13），23%（3/13），和 60%（3/5）。7T-OSU 肿瘤在 RIZ 基因 Pro-704 上有 LOH，但保留了 4 号内显子中的 CA 序列的异源性，又表明其断裂点在 RIZ 基因内部。由于 RIZ 基因位于常见的 LOH 位点附近，同时在 RIZ 基因中发现有断裂点，提示 RIZ 基因在家族性和自发性结肠癌中是 1 号染色体短臂 36 位异常的潜在目的基因。

检查 RIZ 基因编码序列发现在其 8 号外显子中有高突变的 A8 和 A9 多聚体。在 37.5%（9/24）的 MSI 阳性 HNPCC 肿瘤中发现有 A8（1 个）和 A9（8 个）多聚体位点错位突变。另外在几个 MSI 阳性细胞株中有 6 个（HCT116, LoVo, LS411N, LS174T, MDAH2774, 和 AN3CA）有 A9 多聚体位点错位突变。在 HCT116 和 AN3CA 细胞株中，突变是双重的 / 半双重的，是由于两个染色体各体

都参与的结果。在 23 个 MSI 阴性肿瘤中未发现有类似的错位突变, 表明这些位点只是在 MSI 阳性肿瘤中是突变热点。

RIZ 错位突变在肿瘤生成过程中有选择优势 为探讨在其它基因中类似的单核苷酸重复序列是否有相似的突变比例, 我们分析了 PMS2 和 DNA 聚合酶中的 A8 多聚体。这些 A8 多聚体在 MSI 阳性肿瘤中没有任何错位突变 ($P=0.0005$)。在 MSI 阳性肿瘤中又进行 RECQL, BLM, 和 KIAA0355 基因 A9 序列分析。在 24 个 MSI 阳性肿瘤中, 只有一个在 RECQL 和 KIAA0355 基因中有一个碱基丢失 ($P=0.005$)。综合起来, 上述数据表明 RIZ 基因错位突变是结肠癌肿瘤生成中特异的进化选择。

细胞株中 RIZ 基因表达水平下降 以前的数据表明 RIZ 基因有两个转录产物, RIZ1 和 RIZ2 (He 等, Supra, 1998)。在多种与 1p36 有关的肿瘤中, RIZ1 表达丧失, 而 RIZ2 表达不变 (He 等, Supra, 1998; Jiang 等, Int. J. Cancer 83:541-547, 1999; Jiang 等, Histol. Histopathol. 15:109-117, 2000)。为确定在有 RIZ 突变的结肠癌中 RIZ 表达是否受到影响, 我们用 RT-PCR 方法来检测 RIZ1 表达 (He 等, Supra, 1998; Jiang 等, Supra, 1999; Jiang 等, Supra, 2000)。由于 RIZ2 全部包含在 RIZ1 中, 因此无法单独测定 RIZ2 表达。我们只能测 RIZ1 和 RIZ1+RIZ2 表达水平。乳腺癌细胞株 MDAMB435S 作为只表达 RIZ2 的对照, 而 MDAMB21 作为表达 RIZ1+RIZ2 的对照 (He 等, Supra, 1998)。

在检测的 11 个 MSI 阳性结肠癌细胞株中有 4 个 RIZ1 表达降低或丧失而 RIZ2 表达存在, 支持以前观察到的 RIZ1 和 RIZ2 表达不平衡与细胞变化有关 (He 等, Supra, 1998; Jiang 等, Supra, 1999; Jiang 等, Supra, 2000)。在 3 个 RIZ1 表达有变化的细胞中, 包括 HCT116, 存在有多聚腺苷酸的错位突变。在这些细胞中, LS411N 只表达 RIZ2, LoVo 中 RIZ1 表达降低, SW48 中 RIZ1 表达降低, 但没有 RIZ 基因多聚腺苷酸的错位突变。整个 RIZ 基因很大, 有 8 Kb, 可能在这个细胞中在 RIZ 基因其它地方, 或在启动子中有基因突变, 影响基因表达。

在突变阳性的 HNPCC 肿瘤冻组织中抽提的 RNA 没有 RIZ1 表达的特别变化, 这可能是由于在抽提肿瘤的 RNA 时受到正常组织的污染。细胞株中滴度试验表明, 即使在 90% 的 RIZ2 中含有 10% 的 RIZ1/RIZ2 RNA, RIZ1 亦会被扩增。

错位突变导致 RIZ 蛋白质表达中断 由于在 RIZ 基因核苷酸 4700 位置的 A9 处的一个碱基丢失, 造成一个融合 RIZ1 和 RIZ2 蛋白缺少 C-端 219 个氨基酸, 但增加了 76 个新氨基酸, 最终表达的 RIZ1 和 RIZ2 蛋白比正常少 157 个氨基酸。为确定突变的染色体各体确实表达这样的突变 RIZ 蛋白质, 在有错位突变的 HCT116 细胞中进行免疫沉淀和免疫转移膜杂交实验。我们首先确定在 DLD1 细胞中表达正常的 RIZ1 和 RIZ2 蛋白。在以前研究的所有肿瘤细胞株中, RIZ1

表达量极低, 很难检测到, 而 RIZ2 量很高(He 等, Supra, 1998; Buyse 等, Supra, 1995; Liu 等, J. Biol. Chem. 272:2984-2991, 1997)。与此同时, 在 DLD1 细胞中检测到 250 KD 的 RIZ2, 而没有看到 RIZ1。相反在 HCT116 细胞中, 而全长的 RIZ2 不存在, 而看到一个 230 KD 的 RIZ2 品种, 与错位突变形成的 RIZ2 大小相一致。虽然此实验并未证明 RIZ1 表达, 但由 RIZ2 的结果可以推测 RIZ1 亦有翻译中断。

RIZ1 在结肠肿瘤细胞株中表达引起细胞周期阻断和细胞凋亡 腺病毒介导的 RIZ1 表达已被证明在非 MSI 阳性的乳腺癌和肝癌中能造成 G2/M 细胞周期阻断和细胞凋亡。我们探讨了腺病毒介导的 RIZ1 表达在 HCT116 和 DLD1 结肠癌细胞株中的作用。免疫转移分析确定在 AdRIZ1 侵染后(MOI=100), 两个细胞株都有全长的 RIZ1 蛋白表达。48 小时后, 两个细胞中处于 G2/M 期的细胞量明显增加。在 AdRIZ1 侵染后 48 小时和 72 小时, HCT116 细胞中有低于 G1 期 DNA 含量的细胞, 显示有细胞凋亡发生(表一)。但在 DLD1 细胞中这类细胞很少, 结果表明 RIZ1 在 HCT116 细胞中引起 G2/M 阻断和细胞凋亡, 而在 DLD1 细胞中只引起 G2/M 阻断。

表一 Table 1. RIZ1 在结肠肿瘤细胞株 JCT116 和 DLD1 中表达引起细胞周期阻断和细胞凋亡。

Cell lines	Apoptosis	G1	S	G2/M
HCT116	25 \pm 1%	-47 \pm 3%	6.1 \pm 7%	20 \pm 8%
DLD1	9.5 \pm 2%	-28 \pm 1%	2.2 \pm 5%	26 \pm 7%

在 AdRIZ1 或 AdNull 病毒侵染 72 小时后, 细胞进行 DNA 组分分析。表中数据为 Ad RIZ1 与 AdNull 侵染的细胞间差异百分比。

AdRIZ1 侵染后比 AdNull 侵染增加的数目以正值表示, 减少数以负值表示。数据为三次实验的平均值和标准误差。

上述结果表明在肿瘤中 RIZ1 功能可能由于至少两种染色体变化而受影响。在 MSI 阳性肿瘤中, 在基因 3' 端的错位突变影响其 C-端 PR 结合位点与 N-端 PR 区域的相互作用。在 MSI 阴性肿瘤中(CIN 途径), RIZ1 PR 区域的突变或丢失亦有同样效果。在一系列检测的肿瘤中, RIZ 基因因 LOH 或错位突变而受影响, 表明二者在与 RIZ 相关的肿瘤生成中起作用。

实施例 2、在 MSI 阳性的胃癌中的 RIZ 多聚核苷酸错位突变

这个实施例表明 RIZ 多聚核苷酸错位突变在 MSI 阳性的胃癌中发生率很高。

为检测 RIZ 在 MSI 阳性肿瘤中的作用, 我们检测了一系列从朝鲜病人身上取得的 MSI 阳性胃癌样品, 并通过 PCR 和 SSCP 以及测序等方法来分析 RIZ 基因 A8 和 A9 的变化。我们亦检测了一系列 MSI 阳性细胞株, 包括结肠癌 (HCT116, SW-48, LoVo, HCT8), 前列腺癌 (DU145), 乳腺癌 (Cal-51), 子宫癌 (AN3CA, SK-UF18B)。以 MSI 阴性的 SW620 作为对照。

在 33 个 MSI 阳性的胃癌中, 有 16 个 (占 48%), 在 13 个 MSI 阳性的细胞株中有 7 个 (占 54, 其中有 HCT116, HCT8, LoVo, LS441N, LSL80, LSI74T, 和 AN3CA) 有 RIZ 突变。这些突变是体细胞突变, 因为在正常的相应细胞中无此突变。除了在 KSI9 细胞中有 A8 突变外, 其他的都是 A9 位点突变。在 70 个 MSI 阴性的胃癌中未发现 RIZ 突变, 表明这些突变具有 MSI 阳性肿瘤特异性。

在 21 例 RIZ 突变中 (16 个胃癌和 7 个细胞株), 4 例 (KS15, KS19, HCT116, AN3CA) 是两个染色体同时突变。KS15 和 HCT116 是同源/半同源突变。KS19 在一个染色体上的 A8 和 A9 中各有一个碱基丢失, 另一个染色体上在 A9 位点有一个碱基丢失。而 AN3CA 在一个染色体上 A9 中有一个碱基丢失, 另一个染色体上在 A9 位点有二个碱基丢失。

为确定在 MSI 阳性胃癌中是否存在基因组丢失, 我们在 16 个有错位突变的胃癌中选 14 个, 在 17 个没有错位突变的胃癌中选 16 个进行 LOH 检测。他们都有正常的相应组织 DNA。RIZ 蛋白由于在基因 8 号外显子的 704 位氨基酸丢失三个碱基造成 704 号脯氨酸的丢失多样性, 这一点可用来检测 LOH。在已经分析过的 6 个肿瘤中发现有 2 例可检测到 LOH, KS07 和 KS20, 二者在一个染色体上有错位突变。一共检查到 4 个胃癌 (KS07, KS15, KS19, KS20) 和两个细胞株 (HCT-116, AN3CA) 有双染色体 RIZ 失活。

此例中检测到的所有 RIZ 错位突变都预测 RIZ 蛋白在其 C-端中断。在序列 3 的 A9 中丢失一个碱基导致在其后 76 位有终止密码, 其蛋白缺少 C-端 219 个氨基酸; 在 A9 中丢失二个碱基导致在其后 5 位有终止密码; 在 A8 位中丢失一个腺嘌呤导致在其后 2 位有终止密码, 其蛋白缺少 C-端 293 个氨基酸。

已知 RIZ1 的 C-端有 PR 区域结合位点, 可能在与 RIZ1 结合中起作用 (Huang 等, J. Biol. Chem. 273:15933-15440, 1998)。因而此 C-端的缺失很可能会严重影响 RIZ1 功能。

观察到的在 MSI 阳性胃癌中高频率 RIZ 错位突变 (48% 肿瘤, 54% 细胞株) 表明 RIZ 失活与 MSI 阳性胃癌转化有关。经常发现在 RIZ 中有双个染色体的错位突变或基因组突变支持这一预测。另外高频率 RIZ 错位突变亦在其他肿瘤, 包括结肠癌 (示实施例 1) 中发现过。

实施例 3、RIZ 表达在体内抑制 MSI 阳性肿瘤发生，并导致肿瘤细胞凋亡

为确定 RIZ 基因治疗在 MSI 阳性治疗中的效果，我们研究了在已经建立的固体肿瘤中，重组 RIZ1 表达对肿瘤生长的影响，同时也研究了 P53 表达的作用。

从 ATCC 取得的 HCT116 结肠癌细胞株带有同源性 RIZ1 突变和野生型 P53 基因。将其培养在 Delbecco 调配的 Eagle 培养基内，添加 10% 的胎牛血清。

无胸腺雌性裸鼠 (Nu/Nu) 由 Jackson 实验室 (Bar Harbor, Maine) 购得。在 100 微升 PSB 中的 2×10^6 HCT116 肿瘤细胞接种到每一个裸鼠上，6 天后肿瘤开始出现，大致直径为 0.5 cm。治疗前，将动物根据肿瘤大小随机分组，每组 5—11 个。已长肿瘤的裸鼠接受肿瘤内或肿瘤外周注射 PBS，或腺病毒悬浮液 (8×10^{10} 病毒颗粒/剂量)，每隔三天注射一次，共注射 4 次。

缺少基因插入的空腺病毒载体 AdNull (由国家肿瘤研究所 Prem Seth 提供)，或表达 RIZ1 的 AdRIZ1，或表达 p53 蛋白的 Adp53，腺病毒载体的制备，扩增，和滴度测定在 293 细胞上进行，见 He 等，Cancer Res. 58:4328-4244, 1998。

每周测量一次肿瘤大小，肿瘤体积计算公式为 $a \times b^2 \times 0.5$ ，其中 a 是长度，b 是宽度，单位为毫米。用 Student T 测试来比较不同治疗组在每天的肿瘤体积。

如图 2 所示，在 PBS，AdNull，Adp53 治疗的小鼠中肿瘤生长迅速，相反用 AdRIZ1 肿瘤的小鼠肿瘤生长却明显减缓。

我们亦确定了在建立的肿瘤中 RIZ1 和 p53 蛋白表达与细胞凋亡的关系。先将 HCT116 细胞接种到裸鼠中，让其生长一个月，再在肿瘤中注射病毒。注射两天后，将肿瘤取出，固定在 10% 福尔马林溶液中，然后制成石蜡切片，进行组织免疫染色来检定 p53 蛋白，RIZ 表达及 DNA 片段的存在。

用于 RIZ 免疫染色的 1637 抗体血清需稀释 1: 400 倍，用于 p53 蛋白染色的兔 AB545 抗血清稀释 1: 200 倍。二级抗体是过氧化物标记的羊抗兔 IgG。检测细胞凋亡时，按试剂盒使用说明用 Apotag 原位杂交检测试剂盒 (Intergen, Purchase, NY) 进行。简单的讲，去石蜡后，组织再充水，用蛋白酶处理，并与 TdT 培养，用 Avidin-过氧化物酶联合试剂和 DAB 染色。切片用 Hematoxylin 反向染色。

如图 3 所示，大部分注射过 p53 或 RIZ1 病毒的肿瘤在注射部位附近分别在细胞核内表达 p53 或 RIZ1 蛋白质。在 AdRIZ1 病毒注射过的肿瘤，而非 PBS，AdNull，或 Adp53 注射过的肿瘤中有很多很强的 Apotag 染色，证明在 RIZ1 表达的细胞中有明显的细胞凋亡发生。

上述结果与在例 1 中体外观察到的结果相一致，这里的动物试验表明用 RIZ 基因表达系统来处理 MSI 阳性肿瘤可抑制肿瘤生长并引发肿瘤细胞凋亡。

01.03.23

所有上述所引述的文献，参考资料，和专利都在此有备案以供参考。

虽然本发明由上述三个实施例证明，读者应明白对上述三个实施例可做各种修改，但不离开本发明的宗旨。本发明只限于发明专利权利要求所涵盖的范围。

参考资料:

- Aaltonen, et al., 1998, N. Engl. J. Med. 338, 1481-1487
- Anderson, 1998, Nature 392:25-30
- Boland, C.R., et al., 1998, Cancer Res. 58, 5248-57.
- Buyse, I. M., et al., 1995, Proc. Natl. Acad. Sci. U.S.A. 92, 4467-4471
- Canzian, F., et al., 1996, Cancer Res. 56, 3331-7
- Chadwick, R. B., et al., 2000, Proc. Natl. Acad. Sci. USA. 97, 2662-2667
- Chang, Somatic Gene Therapy, CRC Press, Boca Raton, FL, 1995
- Claij, et al., 1999, Exp. Cell. Res. 246, 1-10
- Clayman, 2000, Seminars in Oncology 2788, 39-43
- Dai, et al., 1992, Proc. Natl. Acad. Sci. 89, 10892-10895
- Garriga, et al., 1993, Genes Devel. 7, 2097-2109
- He, L., et al., 1998, Cancer Res. 58, 4238-4244
- Huang, S., et al., 1998 J. Biol. Chem. 273, 15933-15940
- Jiang, G.-L., et al., 1999, Int. J. Cancer 83, 541-547
- Jiang, G.-L. and Huang, S. 2000, Histol. Histopathol. 15, 109-117
- Kaplitt and Loewy, Viral Vectors. Gene Therapy and Neuroscience Applications. Acad. Press, San Diego, CA 1995;
- Leygue, E., et al., 1998, Cancer Res. 58, 1348-52.
- Liu, L., et al., 1997, J. Biol. Chem. 272, 2984-2991
- Mao et al., 1994, Proc. Natl. Acad. Sci. USA 91, 9871-9875
- Morishita et al., 1998, cell 54, 831-840
- Morishita et al., 1990, Oncogene, 5, 936-971
- Roth et al., 1999, Oncology, 13S5, 148-153
- Swisher, S. G., et al., 1999, J. Nat. Cancer Ins. 91, 763-71
- Xie, M., et al., 1997, J. Biol. Chem. 272, 26360-26366.

01.07.03

序列表

所列序列均在本申请人的另一篇专利申请 U. S. 08/516, 859(1995 年 8 月 18 日) 中有描述。

SEQ ID NO. 1:

(i) 序列特征:

- (A) 长度: 6171 碱基对
- (B) 类型: 核酸
- (C) 股链: 双
- (D) 结构: 线性

(ii) 特征

- (A) 命名/关键词: CDS
- (B) 位置: 157-5275

(xi) 序列描述: SEQ ID NO. 1:

GCCAAGATGG CGGCGGCGCG GCCGCGGGCG CCAGGGCGAC GGCGGCGGCT GAGGCTCTGG	60
GCTCGCTGAA GCGTTGGCAC GTCGCGCTCT GGGCTCATGT AATCAAAGAA GTTTCTTTGT	120
TGTGTGTATC TTCACAGAAC ACAACAGGAA TTGAAA ATG CAT CAG AAC ACT GAG	174
Met His Gln Asn Thr Glu	5
TCT GTG GCA GCC ACT GAG ACT CTG GCT GAG GTA CCT GAA CAT GTG CTT	222
Ser Val Ala Ala Thr Glu Thr Leu Ala Glu Val Pro Glu His Val Leu	20
CGA GGA CTT CCA GAG GAA GTA AGA CTT TTC CCA TCT GCA GTC GAC AAG	270
Arg Gly Leu Pro Glu Glu Val Arg Leu Phe Pro Ser Ala Val Asp Lys	35
ACT CGG ATT GGT GTC TGG GCT ACT AAA CCA ATT TTA AAA GGG AAA AAG	318
Thr Arg Ile Gly Val Trp Ala Thr Lys Pro Ile Leu Lys Gly Lys Lys	50
TTT GGG CCA TTT GTT GGT GAT AAG AAG AAG AGA TCC CAG GTT AGG AAT	366
Phe Gly Pro Phe Val Gly Asp Lys Lys Lys Arg Ser Gln Val Arg Asn	70
AAT GTG TAC ATG TGG GAG GTC TAC TAC CCA AAT TTG GGG TGG ATG TGC	414
Asn Val Tyr Met Trp Glu Val Tyr Tyr Pro Asn Leu Gly Trp Met Cys	85
ATT GAT GCC ACC GAT CCG GAG AAG GGC AAC TGG CTA CGC TAT GTG AAC	462
Ile Asp Ala Thr Asp Pro Glu Lys Gly Asn Trp Leu Arg Tyr Val Asn	100
TGG GCT TGC TCA GGA GAA GAG CAG AAT TTA TTT CCA CTG GAA ATC AAC	510
Trp Ala Cys Ser Gly Glu Glu Gln Asn Leu Phe Pro Leu Glu Ile Asn	115
AGA GCC ATT TAC TAT AAA ACC TTA AAG CCA ATC GCG CCT GGC GAG GAG	558
Arg Ala Ile Tyr Tyr Lys Thr Leu Lys Pro Ile Ala Pro Gly Glu Glu	130
CTC CTG GTC TGG TAC AAT GGG GAA GAC AAC CCT GAG ATA GCA GCT GCG	606
Leu Leu Val Trp Tyr Asn Gly Glu Asp Asn Pro Glu Ile Ala Ala Ala	150
ATT GAG GAA GAG CGA GCC AGC GCC CGG AGC AAG CGG AGC TCC CCG AAG	654
Ile Glu Glu Glu Arg Ala Ser Ala Arg Ser Lys Arg Ser Ser Pro Lys	165
AGC CGC AGA GGG AAG AAG AAA TCA CAC GAG AAC AAA AAC AAA GGC ATC	702
Ser Arg Arg Gly Lys Lys Lys Ser His Glu Asn Lys Asn Lys Gly Ile	180

AGA ACC CAC CCC ACA CAG CTG AAG GCA AGT GAG CTG GAC TCT ACC TTT Arg Thr His Pro Thr Gln Leu Lys Ala Ser Glu Leu Asp Ser Thr Phe 185 190 195	750
GCA AAC ATG AGG GGC TCT GCA GAA GGT CCA AAA GAA GAG GAT GAG AGG Ala Asn Met Arg Gly Ser Ala Glu Gly Pro Lys Glu Glu Asp Glu Arg 200 205 210	798
CCT TTG GCT TCG GCA CCT GAG CAG CCA GCC CCT CTG CCG GAG GTG GGG Pro Leu Ala Ser Ala Pro Glu Gln Pro Ala Pro Leu Pro Glu Val Gly 215 220 225 230	846
AAT CAA GAT GCA GTT CCA CAG GTG GCC ATC CCT CTC CCT GCC TGC GAG Asn Gln Asp Ala Val Pro Gln Val Ala Ile Pro Leu Pro Ala Cys Glu 235 240 245	894
CCA CAG CCA GAG GTA GAT GGG AAA CAA GAA GTC ACA GAC TGT GAG GTC Pro Gln Pro Glu Val Asp Gly Lys Gln Glu Val Thr Asp Cys Glu Val 250 255 260	942
AAT GAT GTG GAG GAA GAG GAG CTG GAA GAG GAA GAG GAG CTG GAA GAG Asn Asp Val Glu Glu Glu Glu Leu Glu Glu Glu Glu Leu Glu Glu 265 270 275	990
GAG GAG GAG GAG GAG TTG GGA GAA GAT GGG GTA GAA GAA GCA GAC ATG Glu Glu Glu Glu Glu Leu Gly Glu Asp Gly Val Glu Glu Ala Asp Met 1038	
CCG AAT GAA AGC TCT GCG AAA GAG CCG GAG ATC CCG TGT GAA GAA AAG Pro Asn Glu Ser Ser Ala Lys Glu Pro Glu Ile Arg Cys Glu Glu Lys 295 300 305 310	1086
CCA GAA GAC TTA TTA GAA GAG CCA CAG AGC ATG TCG AAT GAA GCT CGT Pro Glu Asp Leu Leu Glu Glu Pro Gln Ser Met Ser Asn Glu Ala Arg 315 320 325	1134
GAA GAC TCT CCA GAC GTG ACC CCT CCT CCC CAC ACT CCC AGA GCT AGA Glu Asp Ser Pro Asp Val Thr Pro Pro Pro His Thr Pro Arg Ala Arg 330 335 340	1182
GAG GAG GCC AAC GGT GAT GTA CTT GAG ACA TTT ATG TTT CCG TGT CAG Glu Glu Ala Asn Gly Asp Val Leu Glu Thr Phe Met Phe Pro Cys Gln 345 350 355	1230
CAC TGT GAA AGA AAA TTT GCA ACG AAG CAG GGG CTA GAG CGT CAC ATG His Cys Glu Arg Lys Phe Ala Thr Lys Gln Gly Leu Glu Arg His Met 360 365 370	1278
CAC ATC CAC ATT TCT ACC ATC AAT CAT GCT TTC AAG TGC AAG TAC TGT His Ile His Ile Ser Thr Ile Asn His Ala Phe Lys Cys Lys Tyr Cys 375 380 385 390	1326
GGG AAA CGG TTT GGC ACA CAG ATT AAC AGG AGG CGG CAT GAA CGG CGC Gly Lys Arg Phe Gly Thr Gln Ile Asn Arg Arg Arg His Glu Arg Arg 395 400 405	1374
CAC GAA ACG GGG TTG AAG AGA AGA CCC AGC ATG ACT TTA CAG TCC TCA His Glu Thr Gly Leu Lys Arg Arg Pro Ser Met Thr Leu Gln Ser Ser 410 415 420	1422
GAG GAT CCA GAT GAT GGC AAG GGG GAA AAT GTT ACT TCT AAA GAT GAG Glu Asp Pro Asp Asp Gly Lys Gly Glu Asn Val Thr Ser Lys Asp Glu 425 430 435	1470
TCA AGT CCA CCT CAA CTC GGG CAA GAC TGT TTG ATA TTG AAC TCA GAG Ser Ser Pro Pro Gln Leu Gly Gln Asp Cys Leu Ile Leu Asn Ser Glu 440 445 450	1518

AAA ACC TCA CAG GAA GTA CTG AAT TCA TCT TTT GTG GAA GAA AAT GGT Lys Thr Ser Gln Glu Val Leu Asn Ser Ser Phe Val Glu Glu Asn Gly 455 460 465 470	1566
GAA GTT AAA GAA CTT CAT CCG TGC AAA TAC TGC AAA AAG GTA TTT GGA Glu Val Lys Glu Leu His Pro Cys Lys Tyr Cys Lys Lys Val Phe Gly 475 480 485	1614
ACT CAC ACC AAT ATG AGA CGA CAT CAG CGT AGA GTT CAT GAG CGC CAC Thr His Thr Asn Met Arg Arg His Arg Arg Val His Glu Arg His 490 495 500	1662
CTG ATT CCC AAA GGT GTC AGG CGA AAA GGA GGA CTT CTG GAA GAG CCA Leu Ile Lys Gly Val Arg Arg Lys Gly Gly Leu Leu Glu Glu Pro 505 510 515	1710
CAG CCA CCA GCA GAG CAG GCT CCA CCC TCC CAG AAT GTC TAT GTC CCA Gln Pro Pro Ala Glu Gln Ala Pro Pro Ser Gln Asn Val Tyr Val Pro 520 525 530	1758
AGC ACA GAG CCA GAG GAG GAA GGG GAA ACA GAT GAC GTG TAC ATC ATG Ser Thr Glu Pro Glu Glu Glu Gly Glu Thr Asp Asp Val Tyr Ile Met 535 540 545 550	1806
GAC ATT TCT AGC AAC ATC TCT GAA AAC CTA AAT TAC TAT ATT GAC GGT Asp Ile Ser Ser Asn Ile Ser Glu Asn Leu Asn Tyr Tyr Ile Asp Gly 555 560 565	1854
AAG ATT CAG ACC AAC AGC AGC ACT AGT AAC TGT GAT GTG ATT GAG ATG Lys Ile Gln Thr Asn Ser Ser Thr Ser Asn Cys Asp Val Ile Glu Met 570 575 580	1902
GAG TCT AAT TCT GCA CAC TTG TAT GGC ATA GAC TGT CTG CTC ACT CCA Glu Ser Asn Ser Ala His Leu Tyr Gly Ile Asp Cys Leu Leu Thr Pro 585 590 595	1950
GTG ACC GTG GAG ATT ACT CAG AAC ATA AAG AGC ACT CAG GTC TCT GTG Val Thr Val Glu Ile Thr Gln Asn Ile Lys Ser Thr Gln Val Ser Val 600 605 610	1998
ACA GAT GAT CTT CTC AAA GAC TCT CCC AGC AGC ACA AAT TGT GAG TCT Thr Asp Asp Leu Leu Lys Asp Ser Pro Ser Thr Asn Cys Glu Ser 615 620 625 630	2046
AAG AAA CGG AGG ACT GCC AGT CCA CCT GTG CTC CCC AAA ATT AAA ACG Lys Lys Arg Arg Thr Ala Ser Pro Pro Val Leu Pro Lys Ile Lys Thr 635 640 645	2094
GAG ACG GAG TCT GAT TCC ACA GCA CCC TCG TGT TCC TTA AGT CTG CCC Glu Thr Glu Ser Asp Ser Thr Ala Pro Ser Cys Ser Leu Ser Leu Pro 650 655 660	2142
CTG AGC ATA TCC ACA GCC GAG GTG GTG TCC TTC CAT AAA GAG AAG GGC Leu Ser Ile Ser Thr Ala Glu Val Val Ser Phe His Lys Glu Lys Gly 665 670 675	2190
GTC TAT TTG TCG TCC AAG CTC AAG CAG CTT CTT CAG ACC CAG GAC AAG Val Tyr Leu Ser Ser Lys Leu Lys Gln Leu Leu Gln Thr Gln Asp Lys 680 685 690	2238
CTG ACC CTT CCT GCA GGG TTT TCA GCA GCT GAG ATT CCT AAG TTA GGT Leu Thr Leu Pro Ala Gly Phe Ser Ala Ala Glu Ile Pro Lys Leu Gly 695 700 705 710	2286
CCC GTG TGC GCG TCT GCT CCT GCA TCC ATG TTG CCC GTG ACC TCT AGT Pro Val Cys Ala Ser Ala Pro Ala Ser Met Leu Pro Val Thr Ser Ser 715 720 725	2334

AGG TTT AAG AGA CGC ACC AGC TCT CCA CCG AGC TCT CCA CAG CAC AGC	2382
Arg Phe Lys Arg Arg Thr Ser Ser Pro Pro Ser Ser Pro Gln His Ser	
730 735 740	
CCT GCC CTT CGA GAC TTC GGG AAA CCA AAT GAT GGG AAA GCA GCA TGG	2430
Pro Ala Leu Arg Asp Phe Gly Lys Pro Asn Asp Gly Lys Ala Ala Trp	
745 750 755	
ACA GAC ACA GTC CTG ACT TCC AAG AAA CCC AAG TTA GAA AGT CGT AGT	2478
Thr Asp Thr Val Leu Thr Ser Lys Lys Pro Lys Leu Glu Ser Arg Ser	
760 765 770	
GAC TCA CCA GCA TGG AGT TTG TCT GGG AGA GAT GAA AGA GAA ACC GGA	2526
Asp Ser Pro Ala Trp Ser Leu Ser Gly Arg Asp Glu Arg Glu Thr Gly	
775 780 785 790	
AGC CCT CCT TGC TTT GAT GAA TAC AAA ATA TCA AAG GAA TGG GCA GCC	2574
Ser Pro Pro Cys Phe Asp Glu Tyr Lys Ile Ser Lys Glu Trp Ala Ala	
795 800 805	
AGT TCT ACT TTC AGC AGT GTG TGC AAC CAA CAG CCA TTG GAT TTA TCC	2622
Ser Ser Thr Phe Ser Ser Val Cys Asn Gln Gln Pro Leu Asp Leu Ser	
810 815 820	
AGC GGG GTC AAA CAG AAG TCA GAG GGC ACA GGC AAG ACT CCA GTC CCA	2670
Ser Gly Val Lys Gln Lys Ser Glu Gly Thr Gly Lys Thr Pro Val Pro	
825 830 835	
TGG GAA TCT GTA TTG GAT CTC AGT GTG CAT AAA AAG CCT TGC GAT TCT	2718
Trp Glu Ser Val Leu Asp Leu Ser Val His Lys Lys Pro Cys Asp Ser	
840 845 850	
GAA GGC AAG GAA TTC AAA GAG AAC CAT TTG GCA CAG CCA GCT GCA AAG	2766
Glu Gly Lys Glu Phe Lys Glu Asn His Leu Ala Gln Pro Ala Ala Lys	
855 860 865 870	
AAG AAA AAA CCA ACC ACC TGT ATG CTT CAA AAG GTT CTT CTC AAT GAG	2814
Lys Lys Lys Pro Thr Thr Cys Met Leu Gln Lys Val Leu Leu Asn Glu	
875 880 885	
TAT AAT GGT GTT AGC TTA CCT ACA GAA ACC ACA CCA GAG GTG ACC AGG	2862
Tyr Asn Gly Val Ser Leu Pro Thr Glu Thr Thr Pro Glu Val Thr Arg	
890 895 900	
AGC CCA AGT CCT TGT AAA TCC CCA GAT ACA CAG CCA GAT CCT GAA CTT	2910
Ser Pro Ser Pro Cys Lys Ser Pro Asp Thr Gln Pro Asp Pro Glu Leu	
905 910 915	
GGT CCT GAC TCA AGT TGC TCA GTC CCC ACT GCT GAG TCT CCA CCT GAA	2958
Gly Pro Asp Ser Ser Cys Ser Val Pro Thr Ala Glu Ser Pro Pro Glu	
920 925 930	
GTT GTT GGC CCT TCC TCA CCC CCT CTC CAG ACA GCC TCC TTA TCC TCC	3006
Val Val Gly Pro Ser Ser Pro Pro Leu Gln Thr Ala Ser Leu Ser Ser	
935 940 945 950	
GGT CAG CTG CCT CCT CTC TTA ACC CCC ACA GAG CCT TCT TCC CCT CCC	3054
Gly Gln Leu Pro Pro Leu Leu Thr Pro Thr Glu Pro Ser Ser Pro Pro	
955 960 965	
CCC TGC CCT CCT GTG TTA ACT GTT GCC ACT CCA CCA CCT CCC CTC CTT	3102
Pro Cys Pro Pro Val Leu Thr Val Ala Thr Pro Pro Pro Pro Leu Leu	
970 975 980	
CCA ACC GTC CCT CTC TCC CAC CCC TCT TCT GAT GCC TCC CCT CAG CAG	3150
Pro Thr Val Pro Leu Ser His Pro Ser Ser Asp Ala Ser Pro Gln Gln	
985 990 995	

TGT CCC TCT CCG TTC TCA AAC ACC ACT GCT CAG TCT CCT CTT CCC ATT Cys Pro Ser Pro Phe Ser Asn Thr Thr Ala Gln Ser Pro Leu Pro Ile 1000 1005 1010	3198
CTC TCC CCA ACA GTG TCT CCC TCT CCC TCT CCC ATT CCT CCT GTA GAG Leu Ser Pro Thr Val Ser Pro Ser Pro Ser Pro Ile Pro Pro Val Glu 1015 1020 1025 1030	3246
CCA CTT ATG TCT GCT GCT TCC CCT GGT CCC CCA ACA CTT TCT TCC TCC Pro Leu Met Ser Ala Ala Ser Pro Gly Pro Pro Thr Leu Ser Ser Ser 1035 1040 1045	3294
TCC TCT TCT TCC TCT TCC TTC CCT TCC TCT TCC TGC TCC TCC ACC TCC Ser Ser Ser Ser Ser Phe Pro Ser Ser Ser Cys Ser Ser Thr Ser 1050 1055 1060	3342
CCC TCC CCA CCC CCT CTT TCA GCA GTG TCA TCT GTG GTT TCC TCT GGG Pro Ser Pro Pro Leu Ser Ala Val Ser Ser Val Val Ser Ser Gly 1065 1070 1075	3390
GAC AAC CTG GAG GCA TCT CTG CCT GCA GTA ACT TTC AAA CAG GAG GAG Asp Asn Leu Glu Ala Ser Leu Pro Ala Val Thr Phe Lys Gln Glu Glu 1080 1085 1090	3438
TCA GAG AGT GAA GGT CTG AAA CCC AAG GAA GAG GCC CCA CCT GCA GGG Ser Glu Ser Glu Gly Leu Lys Pro Lys Glu Glu Ala Pro Pro Ala Gly 1095 1100 1105 1110	3486
GGA CAG AGT GTG GTC CAA GAA ACA TTC AGC AAA AAC TTC ATT TGC AAT Gly Gln Ser Val Val Gln Glu Thr Phe Ser Lys Asn Phe Ile Cys Asn 1115 1120 1125	3534
GTC TGT GAA TCG CCT TTT CTT TCC ATT AAA GAC CTA ACC AAA CAT TTA Val Cys Glu Ser Pro Phe Leu Ser Ile Lys Asp Leu Thr Lys His Leu 1130 1135 1140	3582
TCC GTC CAT GCT GAA GAG TGG CCC TTC AAA TGT GAG TTT TGT GTG CAG Ser Val His Ala Glu Glu Trp Pro Phe Lys Cys Glu Phe Cys Val Gln 1145 1150 1155	3630
CTG TTT AAG GTT AAG ACT GAT CTA TCA GAG CAT CGA TTT CTG CTT CAT Leu Phe Lys Val Lys Thr Asp Leu Ser Glu His Arg Phe Leu Leu His 1160 1165 1170	3678
GGG GTT GGA AAT ATC TTT GTG TGT TCT GTA TGT AAG AAA GAA TTT GCC Gly Val Gly Asn Ile Phe Val Cys Ser Val Cys Lys Lys Glu Phe Ala 1175 1180 1185 1190	3726
TTC TTA TGC AAT CTG CAG CAG CAC CAG CGT GAT CTC CAC CCA GAT GAG Phe Leu Cys Asn Leu Gln His Gln Arg Asp Leu His Pro Asp Glu 1195 1200 1205	3774
GTA TGC ACA CAC CAC GAG TTT GAA AGT GGG ACC CTG AGG CCC CAG AAC Val Cys Thr His Glu Glu Phe Glu Ser Gly Thr Leu Arg Pro Gln Asn 1210 1215 1220	3822
TTC ACA GAC CCC AGC AAG GCC AAT GTT GAG CAT ATG CCA AGT TTG CCA Phe Thr Asp Pro Ser Lys Ala Asn Val Glu His Met Pro Ser Leu Pro 1225 1230 1235	3870
GAA GAG CCT TTA GAA ACT TCT AGA GAG GAG GAG TTA AAT GAT TCC TCT Glu Glu Pro Leu Glu Thr Ser Arg Glu Glu Glu Leu Asn Asp Ser Ser 1240 1245 1250	3918
GAA GAG CTT TAC ACG ACC ATC AAA ATA ATG GCT TCT GGA ATA AAG ACG Glu Glu Leu Tyr Thr Thr Ile Lys Ile Met Ala Ser Gly Ile Lys Thr 1255 1260 1265 1270	3966

AAG GAT CCA GAT GTT CGA CTT GGT CTC AAC CAG CAC TAC CCG AGC TTT Lys Asp Pro Asp Val Arg Leu Gly Leu Asn Gln His Tyr Pro Ser Phe 1275 1280 1285	4014
AAA CCT CCT CCA TTT CAG TAC CAC CAT CGA AAC CCT ATG GGG ATA GGG Lys Pro Pro Pro Phe Gln Tyr His His Arg Asn Pro Met Gly Ile Gly 1290 1295 1300	4062
GTG ACA GCC ACC AAC TTC ACC ACC CAC AAT ATT CCA CAG ACT TTC ACT Val Thr Ala Thr Asn Phe Thr Thr His Asn Ile Pro Gln Thr Phe Thr 1305 1310 1315	4110
ACT GCC ATC CGC TGC ACA AAG TGT GGG AAG GGC GTC GAC AAT ATG CCT Thr Ala Ile Arg Cys Thr Lys Cys Gly Lys Gly Val Asp Asn Met Pro 1320 1325 1330	4158
GAG CTG CAT AAG CAT ATC TTG GCG TGT GCG TCT GCA AGT GAC AAG AAG Glu Leu His Lys His Ile Leu Ala Cys Ala Ser Ala Ser Asp Lys Lys 1335 1340 1345 1350	4206
AGG TAC ACC CCT AAG AAA AAC CCA GTG CCC CTG AAA CAA ACT GTG CAG Arg Tyr Thr Pro Lys Lys Asn Pro Val Pro Leu Lys Gln Thr Val Gln 1355 1360 1365	4254
CCC AAA AAT GGA GTG GTG GTT CTA GAC AAC TCT GGG AAA AAT GCC TTC Pro Lys Asn Gly Val Val Val Leu Asp Asn Ser Gly Lys Asn Ala Phe 1370 1375 1380	4302
AGA CGG ATG GGG CAG CCC AAG AGA CTG AGC TTC AAT GTT GAA CTG GGT Arg Arg Met Gly Gln Pro Lys Arg Leu Ser Phe Asn Val Glu Leu Gly 1385 1390 1395	4350
AAA ATG TCT CCA AAC AAG CTC AAG CTG AGT GCG CTG AAG AAG AAA AAC Lys Met Ser Pro Asn Lys Leu Lys Leu Ser Ala Leu Lys Lys Lys Asn 1400 1405 1410	4398
CAG CTG GTG CAG AAG GCC ATC CTT CAG AAG AAC AGA GCC GCG AAG CAG Gln Leu Val Gln Lys Ala Ile Leu Gln Lys Asn Arg Ala Ala Lys Gln 1415 1420 1425 1430	4446
AAG GCG GAC CTG AGG GAT ACT TCC GAG GCG TCC TCA CAC ATC TGC CCG Lys Ala Asp Leu Arg Asp Thr Ser Glu Ala Ser Ser His Ile Cys Pro 1435 1440 1445	4494
TAC TGT GAC AGG GAG TTC ACA TAC ATT GGC AGC CTG AAT AAG CAT GCC Tyr Cys Asp Arg Glu Phe Thr Tyr Ile Gly Ser Leu Asn Lys His Ala 1450 1455 1460	4542
GCC TTC AGC TGT CCT AAA AAA CCT CTT TCT CCT TCC AAA AGA AAA GTT Ala Phe Ser Cys Pro Lys Lys Pro Leu Ser Pro Ser Lys Arg Lys Val 1465 1470 1475	4590
TCC CAT TCG TCT AAG AAA GGT GGC CAT GCA TCA TCT TCT AGC AGT GAC Ser His Ser Ser Lys Lys Gly Gly His Ala Ser Ser Ser Ser Ser Asp 1480 1485 1490	4638
AGA AAC AGC AGC TGC CAC CCC CGG AGG CGG ACT GCA GAT ACC GAG ATC Arg Asn Ser Ser Cys His Pro Arg Arg Arg Thr Ala Asp Thr Glu Ile 1495 1500 1505 1510	4686
AAG ATG CAG AGC ACG CAG GCA CCC TTG GGC AAG ACC AGA GCT CGG AGT Lys Met Gln Ser Thr Gln Ala Pro Leu Gly Lys Thr Arg Ala Arg Ser 1515 1520 1525	4734
ACA GGC CCC GCC CAG GCC TCA CTG CCC TCC TCG TCC TTC AGA TCC AGA Thr Gly Pro Ala Gln Ala Ser Leu Pro Ser Ser Ser Phe Arg Ser Arg 1530 1535 1540	4782

CAG AAT GTC AAA TTT GCA GCT TCA GTG AAA TCC AAA AAA GCA AGC TCT Gln Asn Val Lys Phe Ala Ala Ser Val Lys Ser Lys Lys Ala Ser Ser 1545 1550 1555	4830
TCA TCC TTG AGG AAT TCC AGT CCC ATA AGA ATG GCC AAA ATT ACT CAC Ser Ser Leu Arg Asn Ser Ser Pro Ile Arg Met Ala Lys Ile Thr His 1560 1565 1570	4878
GTC GAG GGC AAA AAA CCC AAA GCT GTT GCC AAG AGT CAT TCT GCT CAG Val Glu Gly Lys Lys Pro Lys Ala Val Ala Lys Ser His Ser Ala Gln 1575 1580 1585 1590	4926
CTC TCA AGC AAA TCC TCC CGA GGC CTG CAT GTC AGA GTG CAG AAG AGC Leu Ser Ser Lys Ser Ser Arg Gly Leu His Val Arg Val Gln Lys Ser 1595 1600 1605	4974
AAA GCT GTC ATA CAG AGC AAG ACT GCC CTG GCC AGT AAG AGG AGA ACA Lys Ala Val Ile Gln Ser Lys Thr Ala Leu Ala Ser Lys Arg Arg Thr 1610 1615 1620	5022
GAC CGG TTC ATA GTG AAA TCT AGA GAG CGC AGC GGG GGC CCA ATC ACC Asp Arg Phe Ile Val Lys Ser Arg Glu Arg Ser Gly Gly Pro Ile Thr 1625 1630 1635	5070
CGA AGC CTT CAG CTG GCA GCT GCT GCG GAC CTG AGT GAA AGC AGG AGA Arg Ser Leu Gln Leu Ala Ala Ala Asp Leu Ser Glu Ser Arg Arg 1640 1645 1650	5118
GAG GAC AGC AGT GCC AGG CAT GAG CTG AAG GAC TTC AGC TAC AGT CTC Glu Asp Ser Ser Ala Arg His Glu Leu Lys Asp Phe Ser Tyr Ser Leu 1655 1660 1665 1670	5166
CGC CTG GCA TCT CGA TGC GGC TCA TCA ACA GCC TCT TAC ATC ACC AGA Arg Leu Ala Ser Arg Cys Gly Ser Ser Thr Ala Ser Tyr Ile Thr Arg 1675 1680 1685	5214
CAA TGC AGA AAG GTC AAG GCC GCC GCA GCA ACT CCG TTC CAG GGA CCC Gln Cys Arg Lys Val Lys Ala Ala Ala Thr Pro Phe Gln Gly Pro 1690 1695 1700	5262
TTC CTC AAA GAG T AGGCACTCTG TCTGCTCCTT AACAGCACCT GAAGTGACCT Phe Leu Lys Glu 1705	5315
GGAATCAGTG AAGCCAAAGG GACCAGCAGT CTGCCCTGCA GAGAGCACTG ACCTCTCCCA	5375
GTTGTGAGAG TGAGAGAACG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG AGAGAGAGAG	5435
AGAATGAGAA TGTGTGTGTG TGTGTGCTGG TGCATGTGTG TGGTCTTCAA GCCAAGGTCC	5495
CAGCCTCAGG AGCAGGACCT TCCATTTC CGTCATCCTC TGGATGATCC TTGGACGTGG	5555
CCCAGAACCG TGCTCTGTGG TGCAGCCATC CTGCCCCGGA GGGGCATCTC CTTCTATGCA	5615
ATTTTTTTTAA AGAGTTCTT GGCCTGCTT TGTGCTTCTT GAGCTGTCCG TTTGCCACCA	5675
CTGGGACTTG GATCTGGCCC TGAGGGGTGG GGAAGAGGGC CTATCTAAGG ATAACTTTTC	5735
AGAGGTCAAG CTCCCCTTCA TGCCACCCCT CCCCCCTGCC CTCACCGACC TTTCCCCAC	5795
ACTGTCTCTG GGAATCAATA GCAGATAGCA TATAGATCCA TCAGGGTTGA GCCTGAACCT	5855
CGGCCCTAGC ACTAGGAAAT CCCCCTTTTC TCCCTAAGCA ACTGGAGCCG CCAGCTTTCA	5915
AGTCATTTCC TCCTTTGAGG TTCTAGAGTC CGAGAGTCTG CTCCGAAGTC TCTCCTGGGA	5975
ACCCGGGAGC CCTCGCACCC AGGACGCAGA CTCTGTGCCC ATTCTTAGAC CTGAGGTAGA	6035

AGAAGCAGTG TTTTGGGACG ATAGGGTGGA GCGTGCCTA CTTTGTCTCC TCTGGTGGGA 6095
 CCTCTACAT CATTGGCATC TGAACCTTGC AAGTTCGCTG CAAAGAGAAG CAAAGGAAAA 6155
 AAAAAAAAAA AAAAAA 6171

SEQ ID NO.2:

(i) 序列特征:

(A) 长度: 1076 氨基酸

(B) 类型: 氨基酸

(D) 结构: 线性

(ii) 分子类型: 蛋白质

Met His Gln Asn Thr Glu Ser Val Ala Ala Thr Glu Thr Leu Ala Glu
 1 5 10 15
 Val Pro Glu His Val Leu Arg Gly Leu Pro Glu Glu Val Arg Leu Phe
 20 25 30
 Pro Ser Ala Val Asp Lys Thr Arg Ile Gly Val Trp Ala Thr Lys Pro
 35 40 45
 Ile Leu Lys Gly Lys Lys Phe Gly Pro Phe Val Gly Asp Lys Lys Lys
 50 55 60
 Arg Ser Gln Val Arg Asn Asn Val Tyr Met Trp Glu Val Tyr Tyr Pro
 65 70 75 80
 Asn Leu Gly Trp Met Cys Ile Asp Ala Thr Asp Pro Glu Lys Gly Asn
 85 90 95
 Trp Leu Arg Tyr Val Asn Trp Ala Cys Ser Gly Glu Glu Gln Asn Leu
 100 105 110
 Phe Pro Leu Glu Ile Asn Arg Ala Ile Tyr Tyr Lys Thr Leu Lys Pro
 115 120 125
 Ile Ala Pro Gly Glu Glu Leu Leu Val Trp Tyr Asn Gly Glu Asp Asn
 130 135 140
 Pro Glu Ile Ala Ala Ala Ile Glu Glu Glu Arg Ala Ser Ala Arg Ser
 145 150 155 160
 Lys Arg Ser Ser Pro Lys Ser Arg Arg Gly Lys Lys Lys Ser His Glu
 165 170 175
 Asn Lys Asn Lys Gly Ile Arg Thr His Pro Thr Gln Leu Lys Ala Ser
 180 185 190
 Glu Leu Asp Ser Thr Phe Ala Asn Met Arg Gly Ser Ala Glu Gly Pro
 195 200 205
 Lys Glu Glu Asp Glu Arg Pro Leu Ala Ser Ala Pro Glu Gln Pro Ala
 210 215 220
 Pro Leu Pro Glu Val Gly Asn Gln Asp Ala Val Pro Gln Val Ala Ile
 225 230 235 240
 Pro Leu Pro Ala Cys Glu Pro Gln Pro Glu Val Asp Gly Lys Gln Glu
 245 250 255

Val Thr Asp Cys Glu Val Asn Asp Val Glu Glu Glu Glu Leu Glu Glu
 260 265 270
 Glu Glu Glu Leu Glu Glu Glu Glu Glu Glu Leu Gly Glu Asp Gly
 275 280 285
 Val Glu Glu Ala Asp Met Pro Asn Glu Ser Ser Ala Lys Glu Pro Glu
 290 295 300
 Ile Arg Cys Glu Glu Lys Pro Glu Asp Leu Leu Glu Glu Pro Gln Ser
 305 310 315 320
 Met Ser Asn Glu Ala Arg Glu Asp Ser Pro Asp Val Thr Pro Pro Pro
 325 330 335
 His Thr Pro Arg Ala Arg Glu Glu Ala Asn Gly Asp Val Leu Glu Thr
 340 345 350
 Phe Met Phe Pro Cys Gln His Cys Glu Arg Lys Phe Ala Thr Lys Gln
 355 360 365
 Gly Leu Glu Arg His Met His Ile His Ile Ser Thr Ile Asn His Ala
 370 375 380
 Phe Lys Cys Lys Tyr Cys Gly Lys Arg Phe Gly Thr Gln Ile Asn Arg
 385 390 395 400
 Arg Arg His Glu Arg Arg His Glu Thr Gly Leu Lys Arg Arg Pro Ser
 405 410 415
 Met Thr Leu Gln Ser Ser Glu Asp Pro Asp Asp Gly Lys Gly Glu Asn
 420 425 430
 Val Thr Ser Lys Asp Glu Ser Ser Pro Pro Gln Leu Gly Gln Asp Cys
 435 440 445
 Leu Ile Leu Asn Ser Glu Lys Thr Ser Gln Glu Val Leu Asn Ser Ser
 450 455 460
 Phe Val Glu Glu Asn Gly Glu Val Lys Glu Leu His Pro Cys Lys Tyr
 465 470 475 480
 Cys Lys Lys Val Phe Gly Thr His Thr Asn Met Arg Arg His Gln Arg
 485 490 495
 Arg Val His Glu Arg His Leu Ile Pro Lys Gly Val Arg Arg Lys Gly
 500 505 510
 Gly Leu Leu Glu Glu Pro Gln Pro Pro Ala Glu Gln Ala Pro Pro Ser
 515 520 525
 Gln Asn Val Tyr Val Pro Ser Thr Glu Pro Glu Glu Glu Gly Glu Thr
 530 535 540
 Asp Asp Val Tyr Ile Met Asp Ile Ser Ser Asn Ile Ser Glu Asn Leu
 545 550 555 560
 Asn Tyr Tyr Ile Asp Gly Lys Ile Gln Thr Asn Ser Ser Thr Ser Asn
 565 570 575
 Cys Asp Val Ile Glu Met Glu Ser Asn Ser Ala His Leu Tyr Gly Ile
 580 585 590
 Asp Cys Leu Leu Thr Pro Val Thr Val Glu Ile Thr Gln Asn Ile Lys
 595 600 605

10505

Ser Thr Gln Val Ser Val Thr Asp Asp Leu Leu Lys Asp Ser Pro Ser
610 615 620

Ser Thr Asn Cys Glu Ser Lys Lys Arg Arg Thr Ala Ser Pro Pro Val
625 630 635 640

Leu Pro Lys Ile Lys Thr Glu Thr Glu Ser Asp Ser Thr Ala Pro Ser
645 650 655

Cys Ser Leu Ser Leu Pro Leu Ser Ile Ser Thr Ala Glu Val Val Ser
660 665 670

Phe His Lys Glu Lys Gly Val Tyr Leu Ser Ser Lys Leu Lys Gln Leu
675 680 685

Leu Gln Thr Gln Asp Lys Leu Thr Leu Pro Ala Gly Phe Ser Ala Ala
690 695 700

Glu Ile Pro Lys Leu Gly Pro Val Cys Ala Ser Ala Pro Ala Ser Met
705 710 715 720

Leu Pro Val Thr Ser Ser Arg Phe Lys Arg Arg Thr Ser Ser Pro Pro
725 730 735

Ser Ser Pro Gln His Ser Pro Ala Leu Arg Asp Phe Gly Lys Pro Asn
740 745 750

Asp Gly Lys Ala Ala Trp Thr Asp Thr Val Leu Thr Ser Lys Lys Pro
755 760 765

Lys Leu Glu Ser Arg Ser Asp Ser Pro Ala Trp Ser Leu Ser Gly Arg
770 775 780

Asp Glu Arg Glu Thr Gly Ser Pro Pro Cys Phe Asp Glu Tyr Lys Ile
785 790 795 800

Ser Lys Glu Trp Ala Ala Ser Ser Thr Phe Ser Ser Val Cys Asn Gln
805 810 815

Gln Pro Leu Asp Leu Ser Ser Gly Val Lys Gln Lys Ser Glu Gly Thr
820 825 830

Gly Lys Thr Pro Val Pro Trp Glu Ser Val Leu Asp Leu Ser Val His
835 840 845

Lys Lys Pro Cys Asp Ser Glu Gly Lys Glu Phe Lys Glu Asn His Leu
850 855 860

Ala Gln Pro Ala Ala Lys Lys Lys Lys Pro Thr Thr Cys Met Leu Gln
865 870 875 880

Lys Val Leu Leu Asn Glu Tyr Asn Gly Val Ser Leu Pro Thr Glu Thr
885 890 895

Thr Pro Glu Val Thr Arg Ser Pro Ser Pro Cys Lys Ser Pro Asp Thr
900 905 910

Gln Pro Asp Pro Glu Leu Gly Pro Asp Ser Ser Cys Ser Val Pro Thr
915 920 925

Ala Glu Ser Pro Pro Glu Val Val Gly Pro Ser Ser Pro Pro Leu Gln
930 935 940

Thr Ala Ser Leu Ser Ser Gly Gln Leu Pro Pro Leu Leu Thr Pro Thr
945 950 955 960

107

Glu Pro Ser Ser Pro Pro Pro Cys Pro Pro Val Leu Thr Val Ala Thr
965 970 975

Pro Pro Pro Pro Leu Leu Pro Thr Val Pro Leu Ser His Pro Ser Ser
980 985 990

Asp Ala Ser Pro Gln Gln Cys Pro Ser Pro Phe Ser Asn Thr Thr Ala
995 1000 1005

Gln Ser Pro Leu Pro Ile Leu Ser Pro Thr Val Ser Pro Ser Pro Ser
1010 1015 1020

Pro Ile Pro Pro Val Glu Pro Leu Met Ser Ala Ala Ser Pro Gly Pro
1025 1030 1035 1040

Pro Thr Leu Ser Ser Ser Ser Ser Ser Ser Ser Ser Phe Pro Ser Ser
1045 1050 1055

Ser Cys Ser Ser Thr Ser Pro Ser Pro Pro Pro Leu Ser Ala Val Ser
1060 1065 1070

Ser Val Val Ser Ser Gly Asp Asn Leu Glu Ala Ser Leu Pro Ala Val
1075 1080 1085

Thr Phe Lys Gln Glu Glu Ser Glu Ser Glu Gly Leu Lys Pro Lys Glu
1090 1095 1100

Glu Ala Pro Pro Ala Gly Gly Gln Ser Val Val Gln Glu Thr Phe Ser
1105 1110 1115 1120

Lys Asn Phe Ile Cys Asn Val Cys Glu Ser Pro Phe Leu Ser Ile Lys
1125 1130 1135

Asp Leu Thr Lys His Leu Ser Val His Ala Glu Glu Trp Pro Phe Lys
1140 1145 1150

Cys Glu Phe Cys Val Gln Leu Phe Lys Val Lys Thr Asp Leu Ser Glu
1155 1160 1165

His Arg Phe Leu Leu His Gly Val Gly Asn Ile Phe Val Cys Ser Val
1170 1175 1180

Cys Lys Lys Glu Phe Ala Phe Leu Cys Asn Leu Gln Gln His Gln Arg
1185 1190 1195 1200

Asp Leu His Pro Asp Glu Val Cys Thr His His Glu Phe Glu Ser Gly
1205 1210 1215

Thr Leu Arg Pro Gln Asn Phe Thr Asp Pro Ser Lys Ala Asn Val Glu
1220 1225 1230

His Met Pro Ser Leu Pro Glu Glu Pro Leu Glu Thr Ser Arg Glu Glu
1235 1240 1245

Glu Leu Asn Asp Ser Ser Glu Glu Leu Tyr Thr Thr Ile Lys Ile Met
1250 1255 1260

Ala Ser Gly Ile Lys Thr Lys Asp Pro Asp Val Arg Leu Gly Leu Asn
1265 1270 1275 1280

Gln His Tyr Pro Ser Phe Lys Pro Pro Pro Phe Gln Tyr His His Arg
1285 1290 1295

Asn Pro Met Gly Ile Gly Val Thr Ala Thr Asn Phe Thr Thr His Asn
1300 1305 1310

Ile Pro Gln Thr Phe Thr Thr Ala Ile Arg Cys Thr Lys Cys Gly Lys
 1315 1320 1325
 Gly Val Asp Asn Met Pro Glu Leu His Lys His Ile Leu Ala Cys Ala
 1330 1335 1340
 Ser Ala Ser Asp Lys Lys Arg Tyr Thr Pro Lys Lys Asn Pro Val Pro
 1345 1350 1355 1360
 Leu Lys Gln Thr Val Gln Pro Lys Asn Gly Val Val Val Leu Asp Asn
 1365 1370 1375
 Ser Gly Lys Asn Ala Phe Arg Arg Met Gly Gln Pro Lys Arg Leu Ser
 1380 1385 1390
 Phe Asn Val Glu Leu Gly Lys Met Ser Pro Asn Lys Leu Lys Leu Ser
 1395 1400 1405
 Ala Leu Lys Lys Lys Asn Gln Leu Val Gln Lys Ala Ile Leu Gln Lys
 1410 1415 1420
 Asn Arg Ala Ala Lys Gln Lys Ala Asp Leu Arg Asp Thr Ser Glu Ala
 1425 1430 1435 1440
 Ser Ser His Ile Cys Pro Tyr Cys Asp Arg Glu Phe Thr Tyr Ile Gly
 1445 1450 1455
 Ser Leu Asn Lys His Ala Ala Phe Ser Cys Pro Lys Lys Pro Leu Ser
 1460 1465 1470
 Pro Ser Lys Arg Lys Val Ser His Ser Ser Lys Lys Gly Gly His Ala
 1475 1480 1485
 Ser Ser Ser Ser Asp Arg Asn Ser Ser Cys His Pro Arg Arg Arg
 1490 1495 1500
 Thr Ala Asp Thr Glu Ile Lys Met Gln Ser Thr Gln Ala Pro Leu Gly
 1505 1510 1515 1520
 Lys Thr Arg Ala Arg Ser Thr Gly Pro Ala Gln Ala Ser Leu Pro Ser
 1525 1530 1535
 Ser Ser Phe Arg Ser Arg Gln Asn Val Lys Phe Ala Ala Ser Val Lys
 1540 1545 1550
 Ser Lys Lys Ala Ser Ser Ser Ser Leu Arg Asn Ser Ser Pro Ile Arg
 1555 1560 1565
 Met Ala Lys Ile Thr His Val Glu Gly Lys Lys Pro Lys Ala Val Ala
 1570 1575 1580
 Lys Ser His Ser Ala Gln Leu Ser Ser Lys Ser Ser Arg Gly Leu His
 1585 1590 1595 1600
 Val Arg Val Gln Lys Ser Lys Ala Val Ile Gln Ser Lys Thr Ala Leu
 1605 1610 1615
 Ala Ser Lys Arg Arg Thr Asp Arg Phe Ile Val Lys Ser Arg Glu Arg
 1620 1625 1630
 Ser Gly Gly Pro Ile Thr Arg Ser Leu Gln Leu Ala Ala Ala Ala Asp
 1635 1640 1645
 Leu Ser Glu Ser Arg Arg Glu Asp Ser Ser Ala Arg His Glu Leu Lys
 1650 1655 1660

010000

Asp Phe Ser Tyr Ser Leu Arg Leu Ala Ser Arg Cys Gly Ser Ser Thr
 1665 1670 1675 1680
 Ala Ser Tyr Ile Thr Arg Gln Cys Arg Lys Val Lys Ala Ala Ala Ala
 1685 1690 1695
 Thr Pro Phe Gln Gly Pro Phe Leu Lys Glu
 1700 1705

SEQ ID NO.3:

(i) 序列特征:

(A) 长度: 5868 碱基对

(B) 类型: 核酸

(C) 股链: 双

(D) 结构: 线性

(ii) 特征

(A) 命名/关键词: CDS

(B) 位置: 121-5278

GGAAAATTTA TTCCCACTGG AAATCAACAG AGCCATTTC TATAAACTT TAAAGGGTTC	60
ATGTAATCAA AGAAGTTTCT TGTGTGTGTG TATCTTTACA GAACACAACA GGAATTGAAA	120
ATG AAT CAG AAC ACT ACT GAG CCT GTG GCG GCC ACC GAG ACC CTG GCT Met Asn Gln Asn Thr Thr Glu Pro Val Ala Ala Thr Glu Thr Leu Ala	168
1 5 10 15	
GAG GTA CCC GAA CAT GTG CTG CGA GGA CTT CCG GAG GAA GTG AGG CTT Glu Val Pro Glu His Val Leu Arg Gly Leu Pro Glu Glu Val Arg Leu	216
20 25 30	
TTC CCT TCT GCT GTT GAC AAG ACC CGG ATT GGT GTC TGG GCC ACT AAA Phe Pro Ser Ala Val Asp Lys Thr Arg Ile Gly Val Trp Ala Thr Lys	264
35 40 45	
CCA ATT TTA AAA GGG AAA AAA TTT GGG CCA TTT GTT GGT GAT AAG AAA Pro Ile Leu Lys Gly Lys Lys Phe Gly Pro Phe Val Gly Asp Lys Lys	312
50 55 60	
AAA AGA TCT CAG GTT AAG AAT AAT GTA TAC ATG TGG GAG GTG TAT TAC Lys Arg Ser Gln Val Lys Asn Asn Val Tyr Met Trp Glu Val Tyr Tyr	360
65 70 75 80	
CCA AAT TTG GGA TGG ATG TGC ATT GAT GCC ACT GAT CCA GAG AAG GGA Pro Asn Leu Gly Trp Met Cys Ile Asp Ala Thr Asp Pro Glu Lys Gly	408
85 90 95	
AAC TGG CTG CGA TAT GTG AAT TGG GCT TGC TCA GGA GAA GAG CAA AAT Asn Trp Leu Arg Tyr Val Asn Trp Ala Cys Ser Gly Glu Glu Gln Asn	456
100 105 110	
TTA TTC CCA CTG GAA ATC AAC AGA GCC ATT TAC TAT AAA ACT TTA AAG Leu Phe Pro Leu Glu Ile Asn Arg Ala Ile Tyr Tyr Lys Thr Leu Lys	504
115 120 125	
CCA ATC GCG CCG GGC GAG GAG CTC CTG GTC TGG TAC AAT GGG GAA GAC Pro Ile Ala Pro Gly Glu Glu Leu Leu Val Trp Tyr Asn Gly Glu Asp	552
130 135 140	

01030

AAC CCT GAG ATA GCA GCT GCG ATT GAG GAA GAG CGA GCC AGC GCC CGG Asn Pro Glu Ile Ala Ala Ala Ile Glu Glu Glu Arg Ala Ser Ala Arg 145 150 155 160	600
AGC AAG CGG AGC TCC CCC AAG AGC CGG AAA GGG AAG AAA AAA TCC CAG Ser Lys Arg Ser Ser Pro Lys Ser Arg Lys Gly Lys Lys Lys Ser Gln 165 170 175	648
GAA AAT AAA AAC AAA GGA AAC AAA ATC CAA GAC ATA CAA CTG AAG ACA Glu Asn Lys Asn Lys Gly Asn Lys Ile Gln Asp Ile Gln Leu Lys Thr 180 185 190	696
AGT GAG CCA GAT TTC ACC TCT GCA AAT ATG AGA GAT TCT GCA GAA GGT Ser Glu Pro Asp Phe Thr Ser Ala Asn Met Arg Asp Ser Ala Glu Gly 195 200 205	744
CCT AAA GAA GAC GAA GAG AAG CCT TCA GCC TCA GCA CTT GAG CAG CCG Pro Lys Glu Asp Glu Glu Lys Pro Ser Ala Ser Ala Leu Glu Gln Pro 210 215 220	792
GCC ACC CTC CAG GAG GTG GCC AGT CAG GAG GTG CCT CCA GAA CTA GCA Ala Thr Leu Gln Glu Val Ala Ser Gln Glu Val Pro Pro Glu Leu Ala 225 230 235 240	840
ACC CCT GCC CCT GCC TGG GAG CCA CAG CCA GAA CCA GAC GAG CGA TTA Thr Pro Ala Pro Ala Trp Glu Pro Gln Pro Glu Pro Asp Glu Arg Leu 245 250 255	888
GAA GCG GCA GCT TGT GAG GTG AAT GAT TTG GGG GAA GAG GAG GAG GAG Glu Ala Ala Ala Cys Glu Val Asn Asp Leu Gly Glu Glu Glu Glu Glu 260 265 270	936
GAA GAG GAG GAG GAT GAA GAA GAA GAA GAA GAT GAT GAT GAT GAT GAG Glu Glu Glu Glu Asp Glu Glu Glu Glu Glu Glu Asp Asp Asp Asp Glu 275 280 285	984
TTG GAA GAC GAG GGG GAA GAA GAA GCC AGC ATG CCA AAT GAA AAT TCT Leu Glu Asp Glu Gly Glu Glu Glu Ala Ser Met Pro Asn Glu Asn Ser 290 295 300	1032
GTG AAA GAG CCA GAA ATA CGG TGT GAT GAG AAG CCA GAA GAT TTA TTA Val Lys Glu Pro Glu Ile Arg Cys Asp Glu Lys Pro Glu Asp Leu Leu 305 310 315 320	1080
GAG GAA CCA AAA ACA ACT TCA GAA GAA ACT CTT GAA GAC TGC TCA GAG Glu Glu Pro Lys Thr Thr Ser Glu Glu Thr Leu Glu Asp Cys Ser Glu 325 330 335	1128
GTA ACA CCT GCC ATG CAA ATC CCC AGA ACT AAA GAA GAG GCC AAT GGT Val Thr Pro Ala Met Gln Ile Pro Arg Thr Lys Glu Glu Ala Asn Gly 340 345 350	1176
GAT GTA TTT GAA ACG TTT ATG TTT CCG TGT CAA CAT TGT GAA AGG AAG Asp Val Phe Glu Thr Phe Met Phe Pro Cys Gln His Cys Glu Arg Lys 355 360 365	1224
TTT ACA ACC AAA CAG GGG CTT GAG CGT CAC ATG CAT ATC CAT ATA TCC Phe Thr Thr Lys Gln Gly Leu Glu Arg His Met His Ile His Ile Ser 370 375 380	1272
ACC GTC AAT CAT GCT TTC AAA TGC AAG TAC TGT GGG AAA GCC TTT GGC Thr Val Asn His Ala Phe Lys Cys Lys Tyr Cys Gly Lys Ala Phe Gly 385 390 395 400	1320
ACA CAG ATT AAC CGG CGG CGA CAT GAG CGG CGC CAT GAA GCA GGG TTA Thr Gln Ile Asn Arg Arg Arg His Glu Arg Arg His Glu Ala Gly Leu 405 410 415	1368

010000

AAG CGG AAA CCC AGC CAA ACA CTA CAG CCG TCA GAG GAT CTG GCT GAT Lys Arg Lys Pro Ser Gln Thr Leu Gln Pro Ser Glu Asp Leu Ala Asp 420 425 430	1416
GGC AAA GCA TCT GGA GAA AAC GTT GCT TCA AAA GAT GAT TCG AGT CCT Gly Lys Ala Ser Gly Glu Asn Val Ala Ser Lys Asp Asp Ser Ser Pro 435 440 445	1464
CCC AGT CTT GGG CCA GAC TGT CTG ATC ATG AAT TCA GAG AAG GCT TCC Pro Ser Leu Gly Pro Asp Cys Leu Ile Met Asn Ser Glu Lys Ala Ser 450 455 460	1512
CAA GAC ACA ATA AAT TCT TCT GTC GTA GAA GAG AAT GGG GAA GTT AAA Gln Asp Thr Ile Asn Ser Ser Val Val Glu Glu Asn Gly Glu Val Lys 465 470 475 480	1560
GAA CTT CAT CCG TGC AAA TAT TGT AAA AAG GTT TTT GGA ACT CAT ACT Glu Leu His Pro Cys Lys Tyr Cys Lys Lys Val Phe Gly Thr His Thr 485 490 495	1608
AAT ATG AGA CGG CAT CAG CGT AGA GTT CAC GAA CGT CAT CTG ATT CCC Asn Met Arg Arg His Gln Arg Arg Val His Glu Arg His Leu Ile Pro 500 505 510	1656
AAA GGT GTA CGG CGA AAA GGA GGC CTT GAA GAG CCC CAG CCT CCA GCA Lys Gly Val Arg Arg Lys Gly Gly Leu Glu Glu Pro Gln Pro Pro Ala 515 520 525	1704
GAA CAG GCC CAG GCC ACC CAG AAC GTG TAT GTA CCA AGC ACA GAG CCG Glu Gln Ala Gln Ala Thr Gln Asn Val Tyr Val Pro Ser Thr Glu Pro 530 535 540	1752
GAG GAG GAA GGG GAA GCA GAT GAT GTG TAC ATC ATG GAC ATT TCT AGC Glu Glu Glu Gly Glu Ala Asp Asp Val Tyr Ile Met Asp Ile Ser Ser 545 550 555 560	1800
AAT ATC TCT GAA AAC TTA AAT TAC TAT ATT GAT GGT AAA ATT CAA ACT Asn Ile Ser Glu Asn Leu Asn Tyr Tyr Ile Asp Gly Lys Ile Gln Thr 565 570 575	1848
AAT AAC AAC ACT AGT AAC TGT GAT GTG ATT GAG ATG GAG TCT GCT TCG Asn Asn Asn Thr Ser Asn Cys Asp Val Ile Glu Met Glu Ser Ala Ser 580 585 590	1896
GCA GAT TTG TAT GGT ATA AAT TGT CTG CTC ACT CCA GTT ACA GTG GAA Ala Asp Leu Tyr Gly Ile Asn Cys Leu Leu Thr Pro Val Thr Val Glu 595 600 605	1944
ATT ACT CAA AAT ATA AAG ACC ACA CAG GTC CCT GTA ACA GAA GAT CTT Ile Thr Gln Asn Ile Lys Thr Thr Gln Val Pro Val Thr Glu Asp Leu 610 615 620	1992
CCT AAA GAG CCT TTG GGC AGC ACA AAT AGT GAG GCC AAG AAG CGG AGA Pro Lys Glu Pro Leu Gly Ser Thr Asn Ser Glu Ala Lys Lys Arg Arg 625 630 635 640	2040
ACT GCG AGC CCA CCT GCA CTG CCC AAA ATT AAG GCC GAA ACA GAC TCT Thr Ala Ser Pro Pro Ala Leu Pro Lys Ile Lys Ala Glu Thr Asp Ser 645 650 655	2088
GAC CCC ATG GTC CCC TCT TGC TCT TTA AGT CTT CCT CTT AGC ATA TCA Asp Pro Met Val Pro Ser Cys Ser Leu Ser Leu Pro Leu Ser Ile Ser 660 665 670	2136
ACA ACA GAG GCA GTG TCT TTC CAC AAA GAG AAA AGT GTT TAT TTG TCA Thr Thr Glu Ala Val Ser Phe His Lys Glu Lys Ser Val Tyr Leu Ser 675 680 685	2184

TCA AAG CTC AAA CAA CTT CTT CAA ACC CAA GAT AAA CTA ACT CCT CCT Ser Lys Leu Lys Gln Leu Leu Gln Thr Gln Asp Lys Leu Thr Pro Pro 690 695 700	2232
GCA GGG ATT TCA GCA ACT GAA ATA GCT AAA TTA GGT CCT GTT TGT GTG Ala Gly Ile Ser Ala Thr Glu Ile Ala Lys Leu Gly Pro Val Cys Val 705 710 715 720	2280
TCT GCT CCT GCA TCA ATG TTG CCT GTG ACC TCA AGT AGG TTT AAG AGG Ser Ala Pro Ala Ser Met Leu Pro Val Thr Ser Ser Arg Phe Lys Arg 725 730 735	2328
CGG ACC AGC TCT CCT CCC AGT TCT CCA CAG CAC AGT CCT GCC CTT CGA Arg Thr Ser Ser Pro Pro Ser Ser Pro Gln His Ser Pro Ala Leu Arg 740 745 750	2376
GAC TTT GGA AAG CCA AGT GAT GGG AAA GCA GCA TGG ACC GAT GCC GGG Asp Phe Gly Lys Pro Ser Asp Gly Lys Ala Ala Trp Thr Asp Ala Gly 755 760 765	2424
CTG ACT TCC AAA AAA TCC AAA TTA GAA AGT CAC AGC GAC TCA CCA GCA Leu Thr Ser Lys Lys Ser Lys Leu Glu Ser His Ser Asp Ser Pro Ala 770 775 780	2472
TGG AGT TTG TCT GGG AGA GAT GAG AGA GAA ACT GTG AGC CCT CCA TGC Trp Ser Leu Ser Gly Arg Asp Glu Arg Glu Thr Val Ser Pro Pro Cys 785 790 795 800	2520
TTT GAT GAA TAT AAA ATG TCT AAA GAG TGG ACA GCT AGT TCT GCT TTT Phe Asp Glu Tyr Lys Met Ser Lys Glu Trp Thr Ala Ser Ser Ala Phe 805 810 815	2568
AGC AGT GTG TGC AAC CAG CAG CCA CTG GAT TTA TCC AGC GGT GTC AAA Ser Ser Val Cys Asn Gln Gln Pro Leu Asp Leu Ser Ser Gly Val Lys 820 825 830	2616
CAG AAG GCT GAG GGT ACA GGC AAG ACT CCA GTC CAG TGG GAA TCT GTC Gln Lys Ala Glu Gly Thr Gly Lys Thr Pro Val Gln Trp Glu Ser Val 835 840 845	2664
TTA GAT CTC AGT GTG CAT AAA AAG CAT TGT AGT GAC TCT GAA GGC AAG Leu Asp Leu Ser Val His Lys Lys His Cys Ser Asp Ser Glu Gly Lys 850 855 860	2712
GAA TTC AAA GAA AGT CAT TCA GTG CAG CCT ACG TGT AGT GCT GTA AAG Glu Phe Lys Glu Ser His Ser Val Gln Pro Thr Cys Ser Ala Val Lys 865 870 875 880	2760
AAA AGG AAA CCA ACC ACC TGC ATG CTG CAG AAG GTT CTT CTC AAT GAA Lys Arg Lys Pro Thr Thr Cys Met Leu Gln Lys Val Leu Leu Asn Glu 885 890 895	2808
TAT AAT GGC ATC GAT TTA CCT GTA GAA AAC CCT GCA GAT GGG ACC AGG Tyr Asn Gly Ile Asp Leu Pro Val Glu Asn Pro Ala Asp Gly Thr Arg 900 905 910	2856
AGC CCA AGT CCT TGT AAA TCC CTA GAA GCT CAG CCA GAT CCT GAC CTC Ser Pro Ser Pro Cys Lys Ser Leu Glu Ala Gln Pro Asp Pro Asp Leu 915 920 925	2904
GGT CCG GGC TCT GGT TTC CCT GCC CCT ACT GTT GAG TCC ACA CCT GAT Gly Pro Gly Ser Gly Phe Pro Ala Pro Thr Val Glu Ser Thr Pro Asp 930 935 940	2952
GTT TGT CCT TCA TCA CCT GCC CTG CAG ACA CCC TCC CTT TCA TCC GGT Val Cys Pro Ser Ser Pro Ala Leu Gln Thr Pro Ser Leu Ser Ser Gly 945 950 955 960	3000

010000

CAG CTG CCT CCT CTC TTG ATC CCC ACA GAT CCC TCT TCC CCT CCA CCC Gln Leu Pro Pro Leu Leu Ile Pro Thr Asp Pro Ser Ser Pro Pro Pro 965 970 975	3048
TGT CCC CCG GTA TTA ACT GTT GCC ACT CCG CCC CCT CCC CTC CTT CCT Cys Pro Pro Val Leu Thr Val Ala Thr Pro Pro Pro Pro Leu Leu Pro 980 985 990	3096
ACC GTA CCT CTT CCA GCC CCC TCT TCC AGT GCA TCT CCA CAC CCA TGC Thr Val Pro Leu Pro Ala Pro Ser Ser Ala Ser Pro His Pro Cys 995 1000 1005	3144
CCC TCT CCA CTC TCA AAT GCC ACC GCA CAG TCC CCA CTT CCA ATT CTG Pro Ser Pro Leu Ser Asn Ala Thr Ala Gln Ser Pro Leu Pro Ile Leu 1010 1015 1020	3192
TCC CCA ACA GTG TCC CCC TCT CCC TCT CCC ATT CCT CCC GTG GAG CCC Ser Pro Thr Val Ser Pro Ser Pro Ser Pro Ile Pro Pro Val Glu Pro 1025 1030 1035 1040	3240
CTG ATG TCT GCC GCC TCA CCC GGG CCT CCA ACA CTT TCT TCT TCC TCC Leu Met Ser Ala Ala Ser Pro Gly Pro Pro Thr Leu Ser Ser Ser Ser 1045 1050 1055	3288
TCT TCA TCT TCC TCC TCC TCT TCG TTT TCT TCT TCA TCT TCC TCC TCT Ser Ser Ser Ser Ser Ser Ser Ser Phe Ser Ser Ser Ser Ser Ser Ser 1060 1065 1070	3336
TCT CCT TCT CCA CCT CCT CTC TCC GCA ATA TCA TCT GTT GTT TCC TCT Ser Pro Ser Pro Pro Pro Leu Ser Ala Ile Ser Ser Val Val Ser Ser 1075 1080 1085	3384
GGT GAT AAT CTG GAG GCT TCT CTC CCC ATG ATA TCT TTC AAA CAG GAG Gly Asp Asn Leu Glu Ala Ser Leu Pro Met Ile Ser Phe Lys Gln Glu 1090 1095 1100	3432
GAA TTA GAG AAT GAA GGT CTG AAA CCC AGG GAA GAG CCC CAG TCT GCT Glu Leu Glu Asn Glu Gly Leu Lys Pro Arg Glu Glu Pro Gln Ser Ala 1105 1110 1115 1120	3480
GCT GAA CAG GAT GTT GTT GTT CAG GAA ACA TTC AAC AAA AAC TTT GTT Ala Glu Gln Asp Val Val Val Gln Glu Thr Phe Asn Lys Asn Phe Val 1125 1130 1135	3528
TGC AAC GTC TGT GAA TCA CCT TTT CTT TCC ATT AAA GAT CTA ACC AAA Cys Asn Val Cys Glu Ser Pro Phe Leu Ser Ile Lys Asp Leu Thr Lys 1140 1145 1150	3576
CAT TTA TCT ATT CAT GCT GAA GAA TGG CCC TTC AAA TGT GAA TTT TGT His Leu Ser Ile His Ala Glu Glu Trp Pro Phe Lys Cys Glu Phe Cys 1155 1160 1165	3624
GTG CAG CTT TTT AAG GAT AAA ACG GAC TTG TCA GAA CAT CGC TTT TTG Val Gln Leu Phe Lys Asp Lys Thr Asp Leu Ser Glu His Arg Phe Leu 1170 1175 1180	3672
CTT CAT GGA GTT GGG AAT ATC TTT GTG TGT TCT GTT TGT AAA AAA GAA Leu His Gly Val Gly Asn Ile Phe Val Cys Ser Val Cys Lys Lys Glu 1185 1190 1195 1200	3720
TTT GCT TTT TTG TGC AAT TTG CAG CAG CAC CAG CGA GAT CTC CAC CCA Phe Ala Phe Leu Cys Asn Leu Gln Gln His Gln Arg Asp Leu His Pro 1205 1210 1215	3768
GAT AAG GTG TGC ACA CAT CAC GAG TTT GAA AGC GGG ACT CTG AGG CCC Asp Lys Val Cys Thr His His Glu Phe Glu Ser Gly Thr Leu Arg Pro 1220 1225 1230	3816

010000

CAG AAC TTT ACA GAT CCC AGC AAG GCC CAT GTA GAG CAT ATG CAG AGC Gln Asn Phe Thr Asp Pro Ser Lys Ala His Val Glu His Met Gln Ser 1235 1240 1245	3864
TTG CCA GAA GAT CCT TTA GAA ACT TCT AAA GAA GAA GAG GAG TTA AAT Leu Pro Glu Asp Pro Leu Glu Thr Ser Lys Glu Glu Glu Glu Leu Asn 1250 1255 1260	3912
GAT TCC TCT GAA GAG CTT TAC ACG ACT ATA AAA ATA ATG GCT TCT GGA Asp Ser Ser Glu Glu Leu Tyr Thr Thr Ile Lys Ile Met Ala Ser Gly 1265 1270 1275 1280	3960
ATA AAG ACA AAA GAT CCA GAT GTT CGA TTG GGC CTC AAT CAG CAT TAC Ile Lys Thr Lys Asp Pro Asp Val Arg Leu Gly Leu Asn Gln His Tyr 1285 1290 1295	4008
CCA AGC TTT AAA CCA CCT CCA TTT CAG TAC CAT CAC CGT AAC CCC ATG Pro Ser Phe Lys Pro Pro Phe Gln Tyr His His Arg Asn Pro Met 1300 1305 1310	4056
GGG ATT GGT GTG ACA GCC ACA AAT TTC ACT ACA CAC AAT ATT CCA CAG Gly Ile Gly Val Thr Ala Thr Asn Phe Thr Thr His Asn Ile Pro Gln 1315 1320 1325	4104
ACT TTC ACT ACC GCC ATT CGC TGC ACA AAG TGT GGA AAA GGT GTC GAC Thr Phe Thr Thr Ala Ile Arg Cys Thr Lys Cys Gly Lys Gly Val Asp 1330 1335 1340	4152
AAT ATG CCG GAG TTG CAC AAA CAT ATC CTG GCT TGT GCT TCT GCA AGT Asn Met Pro Glu Leu His Lys His Ile Leu Ala Cys Ala Ser Ala Ser 1345 1350 1355 1360	4200
GAC AAG AAG AGG TAC ACG CCT AAG AAA AAC CCA GTA CCA TTA AAA CAA Asp Lys Lys Arg Tyr Thr Pro Lys Lys Asn Pro Val Pro Leu Lys Gln 1365 1370 1375	4248
ACT GTG CAA CCC AAA AAT GGC GTG GTG GTT TTA GAT AAC TCT GGG AAA Thr Val Gln Pro Lys Asn Gly Val Val Val Leu Asp Asn Ser Gly Lys 1380 1385 1390	4296
AAT GCC TTC CGA CGA ATG GGA CAG CCC AAA AGG CTT AAC TTT AGT GTT Asn Ala Phe Arg Arg Met Gly Gln Pro Lys Arg Leu Asn Phe Ser Val 1395 1400 1405	4344
GAG CTC AGC AAA ATG TCG TCG AAT AAG CTC AAA TTA AAT GCA TTG AAG Glu Leu Ser Lys Met Ser Ser Asn Lys Leu Lys Leu Asn Ala Leu Lys 1410 1415 1420	4392
AAA AAA AAT CAG CTA GTA CAG AAA GCA ATT CTT CAG AAA AAC AAA TCT Lys Lys Asn Gln Leu Val Gln Lys Ala Ile Leu Gln Lys Asn Lys Ser 1425 1430 1435 1440	4440
GCA AAG CAG AAG GCC GAC TTG AAA AAT GCT TGT GAG TCA TCC TCT CAC Ala Lys Gln Lys Ala Asp Leu Lys Asn Ala Cys Glu Ser Ser Ser His 1445 1450 1455	4488
ATC TGC CCT TAC TGT AAT CGA GAG TTC ACT TAC ATT GGA AGC CTG AAT Ile Cys Pro Tyr Cys Asn Arg Glu Phe Thr Tyr Ile Gly Ser Leu Asn 1460 1465 1470	4536
AAA CAC GCC GCC TTC AGC TGT CCC AAA AAA CCC CTT TCT CCT CCC AAA Lys His Ala Ala Phe Ser Cys Pro Lys Lys Pro Leu Ser Pro Pro Lys 1475 1480 1485	4584
AAA AAA GTT TCT CAT TCA TCT AAG AAA GGT GGA CAC TCA TCA CCT GCA Lys Lys Val Ser His Ser Ser Lys Lys Gly Gly His Ser Ser Pro Ala 1490 1495 1500	4632

01/01/00

AGT AGT GAC AAA AAC AGT AAC AGC AAC CAC CGC AGA CGG ACA GCG GAT Ser Ser Asp Lys Asn Ser Asn Ser Asn His Arg Arg Arg Thr Ala Asp 1505 1510 1515 1520	4680
GCG GAG ATT AAA ATG CAA AGC ATG CAG ACT CCG TTG GGC AAG ACC AGA Ala Glu Ile Lys Met Gln Ser Met Gln Thr Pro Leu Gly Lys Thr Arg 1525 1530 1535	4728
GCC CGC AGC TCA GGC CCC ACC CAA GTC CCA CTT CCC TCC TCA TCC TTC Ala Arg Ser Ser Gly Pro Thr Gln Val Pro Leu Pro Ser Ser Ser Phe 1540 1545 1550	4776
AGG TCC AAG CAG AAC GTC AAG TTT GCA GCT TCG GTG AAA TCC AAA AAA Arg Ser Lys Gln Asn Val Lys Phe Ala Ala Ser Val Lys Ser Lys Lys 1555 1560 1565	4824
CCA AGC TCC TCC TCT TTA AGG AAC TCC AGC CCG ATA AGA ATG GCC AAA Pro Ser Ser Ser Ser Leu Arg Asn Ser Ser Pro Ile Arg Met Ala Lys 1570 1575 1580	4872
ATA ACT CAT GTT GAG GGG AAA AAA CCT AAA GCT GTG GCC AAG AAT CAT Ile Thr His Val Glu Gly Lys Lys Pro Lys Ala Val Ala Lys Asn His 1585 1590 1595 1600	4920
TCT GCT CAG CTT TCC AGC AAA ACA TCG CGG AGC CTG CAC GTG AGG GTA Ser Ala Gln Leu Ser Ser Lys Thr Ser Arg Ser Leu His Val Arg Val 1605 1610 1615	4968
CAG AAA AGC AAA GCT GTT TTA CAA AGC AAA TCC ACC TTG GCG AGT AAG Gln Lys Ser Lys Ala Val Leu Gln Ser Lys Ser Thr Leu Ala Ser Lys 1620 1625 1630	5016
AAA AGA ACA GAC CGG TTC AAT ATA AAA TCT AGA GAG CGG AGT GGG GGG Lys Arg Thr Asp Arg Phe Asn Ile Lys Ser Arg Glu Arg Ser Gly Gly 1635 1640 1645	5064
CCA GTC ACC CGG AGC CTT CAG CTG GCA GCT GCT GCT GAC TTG AGT GAG Pro Val Thr Arg Ser Leu Gln Leu Ala Ala Ala Asp Leu Ser Glu 1650 1655 1660	5112
AAC AAG AGA GAG GAC GGC AGC GCC AAG CAG GAG CTG AAG GAC TTC AGC Asn Lys Arg Glu Asp Gly Ser Ala Lys Gln Glu Leu Lys Asp Phe Ser 1665 1670 1675 1680	5160
TAC AGC CTC CGC TTG GCG TCC CGA TGC TCT CCA CCA GCC GCC TCT TAC Tyr Ser Leu Arg Leu Ala Ser Arg Cys Ser Pro Pro Ala Ala Ser Tyr 1685 1690 1695	5208
ATC ACC AGG CAG TAT AGG AAG GTC AAA GCT CCG GCT GCA GCC CAG TTC Ile Thr Arg Gln Tyr Arg Lys Val Lys Ala Pro Ala Ala Ala Gln Phe 1700 1705 1710	5256
CAG GGA CCA TTC TTC AAA GAG T AGACACTCTG GCTGCTCCCT GACAGCACCT Gln Gly Pro Phe Phe Lys Glu 1715	5308
GAAGTGACCT GGAATCAGTG AAGCCAAAGG GACTGGCAGT CTGCCTGCAG GGAGTACCGA	5368
CCTATCCAG TTGTGTGAGG CTGCGAGAGA AAGGGAGTGC ATGTGCGCGC GTGCATGTGT	5428
GCGTGCGTGT GTGTTACCGT GTTCTCGTGC GGGCCGTGAG TGGTCTTCAA ACGAGGGTCC	5488
CGAACCCCGG GGCGGCAGGA AGGGGGCCGA CTCCACGCTG TCCTTTGGGA TGATACTTGG	5548
ATGTCAGCTC TTGGGACCGT GTCTGCAGCC CAGCCTTCCT GTTGGGGTGG GGCCTCTCCT	5608
ACTATGCAAT TTTTCAAGAG CTCCTTGACC CTGCTTTTGG CTTCTTGAGT TGTCTTTTGC	5668

01-03-03

CATTATGGGG ACTTTGGTTT GACCCAGGGG TCAGCCCTTT AGGAAGGCCT TCAGGAGGAG 5728
 GCCGAGTTCC CCTTCAGTAC CACCCCTCTC TCCCCACCTG CCCGCTCCCG GCAACATCTC 5788
 TGGGAATCAA CAGCATATTG ACACGTTGGA GCCGAGCCTG AACATGCCCT GACCCCAGCA 5848
 CATGGAAAAC CCCCTTCCTT 5868

SEQ ID NO. 4:

(i) 序列特征:

(A) 长度: 1719 氨基酸

(B) 类型: 氨基酸

(D) 结构: 线性

(ii) 分子类型: 蛋白质

Met Asn Gln Asn Thr Thr Glu Pro Val Ala Ala Thr Glu Thr Leu Ala
 1 5 10 15
 Glu Val Pro Glu His Val Leu Arg Gly Leu Pro Glu Glu Val Arg Leu
 20 25 30
 Phe Pro Ser Ala Val Asp Lys Thr Arg Ile Gly Val Trp Ala Thr Lys
 35 40 45
 Pro Ile Leu Lys Gly Lys Lys Phe Gly Pro Phe Val Gly Asp Lys Lys
 50 55 60
 Lys Arg Ser Gln Val Lys Asn Asn Val Tyr Met Trp Glu Val Tyr Tyr
 65 70 75 80
 Pro Asn Leu Gly Trp Met Cys Ile Asp Ala Thr Asp Pro Glu Lys Gly
 85 90 95
 Asn Trp Leu Arg Tyr Val Asn Trp Ala Cys Ser Gly Glu Glu Gln Asn
 100 105 110
 Leu Phe Pro Leu Glu Ile Asn Arg Ala Ile Tyr Tyr Lys Thr Leu Lys
 115 120 125
 Pro Ile Ala Pro Gly Glu Glu Leu Leu Val Trp Tyr Asn Gly Glu Asp
 130 135 140
 Asn Pro Glu Ile Ala Ala Ala Ile Glu Glu Glu Arg Ala Ser Ala Arg
 145 150 155 160
 Ser Lys Arg Ser Ser Pro Lys Ser Arg Lys Gly Lys Lys Lys Ser Gln
 165 170 175
 Glu Asn Lys Asn Lys Gly Asn Lys Ile Gln Asp Ile Gln Leu Lys Thr
 180 185 190
 Ser Glu Pro Asp Phe Thr Ser Ala Asn Met Arg Asp Ser Ala Glu Gly
 195 200 205
 Pro Lys Glu Asp Glu Glu Lys Pro Ser Ala Ser Ala Leu Glu Gln Pro
 210 215 220
 Ala Thr Leu Gln Glu Val Ala Ser Gln Glu Val Pro Pro Glu Leu Ala
 225 230 235 240
 Thr Pro Ala Pro Ala Trp Glu Pro Gln Pro Glu Pro Asp Glu Arg Leu
 245 250 255

01-10-20

Glu Ala Ala Ala Cys Glu Val Asn Asp Leu Gly Glu Glu Glu Glu Glu
260 265 270

Glu Glu Glu Glu Asp Glu Glu Glu Glu Glu Asp Asp Asp Asp Glu
275 280 285

Leu Glu Asp Glu Gly Glu Glu Glu Ala Ser Met Pro Asn Glu Asn Ser
290 295 300

Val Lys Glu Pro Glu Ile Arg Cys Asp Glu Lys Pro Glu Asp Leu Leu
305 310 315 320

Glu Glu Pro Lys Thr Thr Ser Glu Glu Thr Leu Glu Asp Cys Ser Glu
325 330 335

Val Thr Pro Ala Met Gln Ile Pro Arg Thr Lys Glu Glu Ala Asn Gly
340 345 350

Asp Val Phe Glu Thr Phe Met Phe Pro Cys Gln His Cys Glu Arg Lys
355 360 365

Phe Thr Thr Lys Gln Gly Leu Glu Arg His Met His Ile His Ile Ser
370 375 380

Thr Val Asn His Ala Phe Lys Cys Lys Tyr Cys Gly Lys Ala Phe Gly
385 390 395 400

Thr Gln Ile Asn Arg Arg Arg His Glu Arg Arg His Glu Ala Gly Leu
405 410 415

Lys Arg Lys Pro Ser Gln Thr Leu Gln Pro Ser Glu Asp Leu Ala Asp
420 425 430

Gly Lys Ala Ser Gly Glu Asn Val Ala Ser Lys Asp Asp Ser Ser Pro
435 440 445

Pro Ser Leu Gly Pro Asp Cys Leu Ile Met Asn Ser Glu Lys Ala Ser
450 455 460

Gln Asp Thr Ile Asn Ser Ser Val Val Glu Glu Asn Gly Glu Val Lys
465 470 475 480

Glu Leu His Pro Cys Lys Tyr Cys Lys Lys Val Phe Gly Thr His Thr
485 490 495

Asn Met Arg Arg His Gln Arg Arg Val His Glu Arg His Leu Ile Pro
500 505 510

Lys Gly Val Arg Arg Lys Gly Gly Leu Glu Glu Pro Gln Pro Pro Ala
515 520 525

Glu Gln Ala Gln Ala Thr Gln Asn Val Tyr Val Pro Ser Thr Glu Pro
530 535 540

Glu Glu Glu Gly Glu Ala Asp Asp Val Tyr Ile Met Asp Ile Ser Ser
545 550 555 560

Asn Ile Ser Glu Asn Leu Asn Tyr Tyr Ile Asp Gly Lys Ile Gln Thr
565 570 575

Asn Asn Asn Thr Ser Asn Cys Asp Val Ile Glu Met Glu Ser Ala Ser
580 585 590

Ala Asp Leu Tyr Gly Ile Asn Cys Leu Leu Thr Pro Val Thr Val Glu
595 600 605

010000

Ile Thr Gln Asn Ile Lys Thr Thr Gln Val Pro Val Thr Glu Asp Leu
610 615 620

Pro Lys Glu Pro Leu Gly Ser Thr Asn Ser Glu Ala Lys Lys Arg Arg
625 630 635 640

Thr Ala Ser Pro Pro Ala Leu Pro Lys Ile Lys Ala Glu Thr Asp Ser
645 650 655

Asp Pro Met Val Pro Ser Cys Ser Leu Ser Leu Pro Leu Ser Ile Ser
660 665 670

Thr Thr Glu Ala Val Ser Phe His Lys Glu Lys Ser Val Tyr Leu Ser
675 680 685

Ser Lys Leu Lys Gln Leu Leu Gln Thr Gln Asp Lys Leu Thr Pro Pro
690 695 700

Ala Gly Ile Ser Ala Thr Glu Ile Ala Lys Leu Gly Pro Val Cys Val
705 710 715 720

Ser Ala Pro Ala Ser Met Leu Pro Val Thr Ser Ser Arg Phe Lys Arg
725 730 735

Arg Thr Ser Ser Pro Pro Ser Ser Pro Gln His Ser Pro Ala Leu Arg
740 745 750

Asp Phe Gly Lys Pro Ser Asp Gly Lys Ala Ala Trp Thr Asp Ala Gly
755 760 765

Leu Thr Ser Lys Lys Ser Lys Leu Glu Ser His Ser Asp Ser Pro Ala
770 775 780

Trp Ser Leu Ser Gly Arg Asp Glu Arg Glu Thr Val Ser Pro Pro Cys
785 790 795 800

Phe Asp Glu Tyr Lys Met Ser Lys Glu Trp Thr Ala Ser Ser Ala Phe
805 810 815

Ser Ser Val Cys Asn Gln Gln Pro Leu Asp Leu Ser Ser Gly Val Lys
820 825 830

Gln Lys Ala Glu Gly Thr Gly Lys Thr Pro Val Gln Trp Glu Ser Val
835 840 845

Leu Asp Leu Ser Val His Lys Lys His Cys Ser Asp Ser Glu Gly Lys
850 855 860

Glu Phe Lys Glu Ser His Ser Val Gln Pro Thr Cys Ser Ala Val Lys
865 870 875 880

Lys Arg Lys Pro Thr Thr Cys Met Leu Gln Lys Val Leu Leu Asn Glu
885 890 895

Tyr Asn Gly Ile Asp Leu Pro Val Glu Asn Pro Ala Asp Gly Thr Arg
900 905 910

Ser Pro Ser Pro Cys Lys Ser Leu Glu Ala Gln Pro Asp Pro Asp Leu
915 920 925

Gly Pro Gly Ser Gly Phe Pro Ala Pro Thr Val Glu Ser Thr Pro Asp
930 935 940

Val Cys Pro Ser Ser Pro Ala Leu Gln Thr Pro Ser Leu Ser Ser Gly
945 950 955 960

01/00/00

Gln Leu Pro Pro Leu Leu Ile Pro Thr Asp Pro Ser Ser Pro Pro Pro
965 970 975

Cys Pro Pro Val Leu Thr Val Ala Thr Pro Pro Pro Pro Leu Leu Pro
980 985 990

Thr Val Pro Leu Pro Ala Pro Ser Ser Ser Ala Ser Pro His Pro Cys
995 1000 1005

Pro Ser Pro Leu Ser Asn Ala Thr Ala Gln Ser Pro Leu Pro Ile Leu
1010 1015 1020

Ser Pro Thr Val Ser Pro Ser Pro Ser Pro Ile Pro Pro Val Glu Pro
1025 1030 1035 1040

Leu Met Ser Ala Ala Ser Pro Gly Pro Pro Thr Leu Ser Ser Ser Ser
1045 1050 1055

Ser Ser Ser Ser Ser Ser Ser Ser Phe Ser Ser Ser Ser Ser Ser Ser
1060 1065 1070

Ser Pro Ser Pro Pro Pro Leu Ser Ala Ile Ser Ser Val Val Ser Ser
1075 1080 1085

Gly Asp Asn Leu Glu Ala Ser Leu Pro Met Ile Ser Phe Lys Gln Glu
1090 1095 1100

Glu Leu Glu Asn Glu Gly Leu Lys Pro Arg Glu Glu Pro Gln Ser Ala
1105 1110 1115 1120

Ala Glu Gln Asp Val Val Val Gln Glu Thr Phe Asn Lys Asn Phe Val
1125 1130 1135

Cys Asn Val Cys Glu Ser Pro Phe Leu Ser Ile Lys Asp Leu Thr Lys
1140 1145 1150

His Leu Ser Ile His Ala Glu Glu Trp Pro Phe Lys Cys Glu Phe Cys
1155 1160 1165

Val Gln Leu Phe Lys Asp Lys Thr Asp Leu Ser Glu His Arg Phe Leu
1170 1175 1180

Leu His Gly Val Gly Asn Ile Phe Val Cys Ser Val Cys Lys Lys Glu
1185 1190 1195 1200

Phe Ala Phe Leu Cys Asn Leu Gln Gln His Gln Arg Asp Leu His Pro
1205 1210 1215

Asp Lys Val Cys Thr His His Glu Phe Glu Ser Gly Thr Leu Arg Pro
1220 1225 1230

Gln Asn Phe Thr Asp Pro Ser Lys Ala His Val Glu His Met Gln Ser
1235 1240 1245

Leu Pro Glu Asp Pro Leu Glu Thr Ser Lys Glu Glu Glu Glu Leu Asn
1250 1255 1260

Asp Ser Ser Glu Glu Leu Tyr Thr Thr Ile Lys Ile Met Ala Ser Gly
1265 1270 1275 1280

Ile Lys Thr Lys Asp Pro Asp Val Arg Leu Gly Leu Asn Gln His Tyr
1285 1290 1295

Pro Ser Phe Lys Pro Pro Pro Phe Gln Tyr His His Arg Asn Pro Met
1300 1305 1310

010000

Gly Ile Gly Val Thr Ala Thr Asn Phe Thr Thr His Asn Ile Pro Gln
1315 1320 1325

Thr Phe Thr Thr Ala Ile Arg Cys Thr Lys Cys Gly Lys Gly Val Asp
1330 1335 1340

Asn Met Pro Glu Leu His Lys His Ile Leu Ala Cys Ala Ser Ala Ser
1345 1350 1355 1360

Asp Lys Lys Arg Tyr Thr Pro Lys Lys Asn Pro Val Pro Leu Lys Gln
1365 1370 1375

Thr Val Gln Pro Lys Asn Gly Val Val Val Leu Asp Asn Ser Gly Lys
1380 1385 1390

Asn Ala Phe Arg Arg Met Gly Gln Pro Lys Arg Leu Asn Phe Ser Val
1395 1400 1405

Glu Leu Ser Lys Met Ser Ser Asn Lys Leu Lys Leu Asn Ala Leu Lys
1410 1415 1420

Lys Lys Asn Gln Leu Val Gln Lys Ala Ile Leu Gln Lys Asn Lys Ser
1425 1430 1435 1440

Ala Lys Gln Lys Ala Asp Leu Lys Asn Ala Cys Glu Ser Ser Ser His
1445 1450 1455

Ile Cys Pro Tyr Cys Asn Arg Glu Phe Thr Tyr Ile Gly Ser Leu Asn
1460 1465 1470

Lys His Ala Ala Phe Ser Cys Pro Lys Lys Pro Leu Ser Pro Pro Lys
1475 1480 1485

Lys Lys Val Ser His Ser Ser Lys Lys Gly Gly His Ser Ser Pro Ala
1490 1495 1500

Ser Ser Asp Lys Asn Ser Asn Ser Asn His Arg Arg Arg Thr Ala Asp
1505 1510 1515 1520

Ala Glu Ile Lys Met Gln Ser Met Gln Thr Pro Leu Gly Lys Thr Arg
1525 1530 1535

Ala Arg Ser Ser Gly Pro Thr Gln Val Pro Leu Pro Ser Ser Ser Phe
1540 1545 1550

Arg Ser Lys Gln Asn Val Lys Phe Ala Ala Ser Val Lys Ser Lys Lys
1555 1560 1565

Pro Ser Ser Ser Ser Leu Arg Asn Ser Ser Pro Ile Arg Met Ala Lys
1570 1575 1580

Ile Thr His Val Glu Gly Lys Lys Pro Lys Ala Val Ala Lys Asn His
1585 1590 1595 1600

Ser Ala Gln Leu Ser Ser Lys Thr Ser Arg Ser Leu His Val Arg Val
1605 1610 1615

Gln Lys Ser Lys Ala Val Leu Gln Ser Lys Ser Thr Leu Ala Ser Lys
1620 1625 1630

Lys Arg Thr Asp Arg Phe Asn Ile Lys Ser Arg Glu Arg Ser Gly Gly
1635 1640 1645

Pro Val Thr Arg Ser Leu Gln Leu Ala Ala Ala Ala Asp Leu Ser Glu
1650 1655 1660

01.03.23

Asn Lys Arg Glu Asp Gly Ser Ala Lys Gln Glu Leu Lys Asp Phe Ser
1665 1670 1675 1680
Tyr Ser Leu Arg Leu Ala Ser Arg Cys Ser Pro Pro Ala Ala Ser Tyr
1685 1690 1695
Ile Thr Arg Gln Tyr Arg Lys Val Lys Ala Pro Ala Ala Ala Gln Phe
1700 1705 1710
Gln Gly Pro Phe Phe Lys Glu
1715

SEQ ID NO.100:

(i) 序列特征:

(A) 长度: 115 氨基酸

(B) 类型: 氨基酸

(D) 结构: 线性

```
Met Pro Gly Ala Gly Leu Gly Ile Trp Thr Lys Arg Lys Ile Glu Val
1          5          10          15
Gly Glu Lys Phe Gly Pro Tyr Val Gly Glu Gln Arg Ser Asn Leu Lys
20          25          30
Asp Pro Ser Tyr Gly Trp Glu Ile Leu Asp Glu Phe Tyr Asn Val Lys
35          40          45
Phe Cys Ile Asp Ala Ser Gln Pro Asp Val Gly Ser Trp Leu Lys Tyr
50          55          60
Ile Arg Phe Ala Gly Cys Tyr Asp Gln His Asn Leu Val Ala Cys Gln
65          70          75          80
Ile Asn Asp Gln Ile Phe Tyr Arg Val Val Ala Asp Ile Ala Pro Gly
85          90          95
Glu Glu Leu Leu Leu Phe Met Lys Ser Glu Asp Tyr Pro His Glu Thr
100          105          110
Met Ala Pro
115
```

SEQ ID NO.101:

(i) 序列特征:

(A) 长度: 110 氨基酸

(B) 类型: 氨基酸

(D) 结构: 线性

```
Asp Glu Leu His Gly Asn Val Leu Ile Ala Val Thr Gln Ile Ala Leu
1          5          10          15
Gly Arg Thr Ile Gly Val Ile Asp Lys Ala Thr Pro Asn Asp Ser Asn
20          25          30
Ala Leu Leu Ile Leu Asn Leu Ile Lys Glu Ala Asp Asp Gly Glu Asp
35          40          45
Ala Asn Ile Cys Met Arg Gln Glu Asp Arg Lys Thr Phe Leu Gln Thr
50          55          60
Ser Lys Ile Ile Asn Ile Gly Glu Arg Leu Leu Leu Gln Arg Leu Ser
65          70          75          80
Glu Glu Glu Cys Asp Glu Glu Asp Gln Asp Asp Leu Glu Asn Leu Ile
85          90          95
Leu Leu Lys Asp Glu Asp Arg Pro Asp Ser Thr Gln Ser Cys
100          105          110
```

SEQ ID NO.102:

(i) 序列特征:

(A) 长度: 21 氨基酸

(B) 类型: 氨基酸

(D) 结构: 线性

```
Glu Trp Gly Pro Val Thr Arg Ser Leu Gln Arg Ser Thr Lys Gln Glu
1          5          10          15
Leu Lys Asp Leu Gln
20
```

SEQ ID NO.103:

(i) 序列特征:

(A) 长度: 22 氨基酸

(B) 类型: 氨基酸

(D) 结构: 线性

```
Gly Ala Glu Glu Thr Ala Ala Ala Ala Ala Leu Gly Ala Leu Arg
1          5          10          15
Leu Gly Arg Arg Ala Arg
20
```

SEQ ID NO.104:

(i) 序列特征:

(A) 长度: 391 碱基对

(B) 类型: 核酸

(C) 股链: 单

(D) 结构: 线性

(ii) 特征

(A) 命名/关键词: CDS

(B) 位置: 1-391

```
GTG TAC TAC CCA AAT TTG GGG TGG ATG TGC ATT GAT GCC ACT GAT CCG      48
Val Tyr Tyr Pro Asn Leu Gly Trp Met Cys Ile Asp Ala Thr Asp Pro
1          5          10          15

GAG AAG GGC AAC TGG CTC CGC TAT GTG AAC TGG GCT TGC TCA GGA GAA      96
Glu Lys Gly Asn Trp Leu Arg Tyr Val Asn Trp Ala Cys Ser Gly Glu
20          25          30

GAA CAG AAT TTA TTT CCA CTG GAA ATC AAC AGA GCC ATT TAC TAT AAA      144
Glu Gln Asn Leu Phe Pro Leu Glu Ile Asn Arg Ala Ile Tyr Tyr Lys
35          40          45

ACC TTA AAG CCA ATC GCG CCT GGC GAG GAG CTC CTG GTC TGG TAC AAT      192
Thr Leu Lys Pro Ile Ala Pro Gly Glu Glu Leu Leu Val Trp Tyr Asn
50          55          60
```

GGG GAA GAC AAC CCC GAG ATA GCA GCT GCG ATT GAG GAA GAG CGA GCC	240
Gly Glu Asp Asn Pro Glu Ile Ala Ala Ala Ile Glu Glu Glu Arg Ala	
65 70 75 80	
AGC GCC CGG AGC AAG CGG AGC TCC CCG AAG AGC CGG AGA GGG AAG AAG	288
Ser Ala Arg Ser Lys Arg Ser Ser Pro Lys Ser Arg Arg Gly Lys Lys	
85 90 95	
AAA TCA CAG GAG AAT AAA AAC AAA GGC ATC AGA ACC CAG GCT GCA GCG	336
Lys Ser Gln Glu Asn Lys Asn Lys Gly Ile Arg Thr Gln Ala Ala Ala	
100 105 110	
CGG AAG GCG AGC GAG CTG GAC TCC ACC TCT GCA AAC ATG AGG GGC TCT	384
Arg Lys Ala Ser Glu Leu Asp Ser Thr Ser Ala Asn Met Arg Gly Ser	
115 120 125	
GCA GAA G	391
Ala Glu	
130	

SEQ ID NO.105:

(i) 序列特征:

(A) 长度: 130 氨基酸

(B) 类型: 氨基酸

(D) 结构: 线性

(ii) 分子类型: 蛋白质

Val Tyr Tyr Pro Asn Leu Gly Trp Met Cys Ile Asp Ala Thr Asp Pro	
1 5 10 15	
Glu Lys Gly Asn Trp Leu Arg Tyr Val Asn Trp Ala Cys Ser Gly Glu	
20 25 30	
Glu Gln Asn Leu Phe Pro Leu Glu Ile Asn Arg Ala Ile Tyr Tyr Lys	
35 40 45	
Thr Leu Lys Pro Ile Ala Pro Gly Glu Glu Leu Leu Val Trp Tyr Asn	
50 55 60	
Gly Glu Asp Asn Pro Glu Ile Ala Ala Ala Ile Glu Glu Glu Arg Ala	
65 70 75 80	
Ser Ala Arg Ser Lys Arg Ser Ser Pro Lys Ser Arg Arg Gly Lys Lys	
85 90 95	
Lys Ser Gln Glu Asn Lys Asn Lys Gly Ile Arg Thr Gln Ala Ala Ala	
100 105 110	
Arg Lys Ala Ser Glu Leu Asp Ser Thr Ser Ala Asn Met Arg Gly Ser	
115 120 125	
Ala Glu	
130	

SEQ ID NO.106:

(i) 序列特征:

(A) 长度: 20 碱基对

(B) 类型: 核酸

(C) 股链: 单

(D) 结构: 线性

(xi) 序列描述: SEQ ID NO.106:

CGCAGGAGCT CCTGGTCTGG

说明书附图

hR1Z	60	ATETLAEPVE HVLRLPEEV RLFPsAVDKT RIGVWATKPI LKCKKFGPFV	60
rR1Z	59	ATETLAEPVE HVLRLPEEV RLFPsAVDKT RIGVWATKPI LKCKKFGPFV	59
一致性	60	ATETLAEPVE HVLRLPEEV RLFPsAVDKT RIGVWATKPI LKCKKFGPFV	60
hR1Z	120	NNVYWEVY PNLGWMCIDA TOPEKGNMLR YVNWACSGEE QNLFPLEINR	120
rR1Z	119	NNVYWEVY PNLGWMCIDA TOPEKGNMLR YVNWACSGEE QNLFPLEINR	119
一致性	120	NNVYWEVY PNLGWMCIDA TOPEKGNMLR YVNWACSGEE QNLFPLEINR	120
hR1Z	180	AIYYKTLKPI APGEELLVY NGEDNPEIAA AIEEERASAR SKRSSPKSRK	180
rR1Z	179	AIYYKTLKPI APGEELLVY NGEDNPEIAA AIEEERASAR SKRSSPKSRK	179
一致性	180	AIYYKTLKPI APGEELLVY NGEDNPEIAA AIEEERASAR SKRSSPKSRK	180
hR1Z	240	KQXKIDQOL KISEDTISA NMRISAEGPK EDEBPEASA EQPATLIEV ASQEVHPEIA	240
rR1Z	239	KQIRTHPTQOL KISEDTISA NMRISAEGPK EDEBPEASA EQPATLIEV ASQEVHPEIA	239
一致性	240	KQXKIDQOL KISEDTISA NMRISAEGPK EDEBPEASA EQPATLIEV ASQEVHPEIA	240
hR1Z	300	TPPAEPQP EDERLEAA CEVNDLCEE EEEEEEE EEDDDDDDE DECEEAAMP	300
rR1Z	295	TPPAEPQP EDERLEAA CEVNDLCEE EEEEEEE EEDDDDDDE DECEEAAMP	295
一致性	300	TPPAEPQP EDERLEAA CEVNDLCEE EEEEEEE EEDDDDDDE DECEEAAMP	300
hR1Z	360	NESSKEPEI RCIEKPEDLL EEFKTTSEET LEDISEVTFA MQIPRTKEEA NCDVETFMF	360
rR1Z	355	NESSKEPEI RCIEKPEDLL EEFKTTSEET LEDISEVTFA MQIPRTKEEA NCDVETFMF	355
一致性	360	NESSKEPEI RCIEKPEDLL EEFKTTSEET LEDISEVTFA MQIPRTKEEA NCDVETFMF	360

420
415
420

PCQHCERKFT	TKQGLERHMH	IHIST	DNHAF	KCKYCGK	FC	TQINRRRRHER	RHE	CLKR	P
PCQHCERKFA	TKQGLERHMH	IHIST	DNHAF	KCKYCGK	FC	TQINRRRRHER	RHE	CLKR	P
PCQHCERKE	TKQGLERHMH	IHIST	DNHAF	KCKYCGK	FC	TQINRRRRHER	RHE	CLKR	P

—致性

480
473
480

STLQSEDP	ADCKASGENV	SKDESSPP	LC	DCLI	NS	EK	SD	NS	SV	EENCEVK
STLQSEDP	ADCKASGENV	SKDESSPP	LC	DCLI	NS	EK	SD	NS	SV	EENCEVK
STLQSED	ADCKASGENV	SKDESSPP	LC	DCLI	NS	EK	SD	NS	SV	EENCEVK

—致性

539
533
540

ELHPCKYCKK	VFGTHTNMRR	HQRRVHERHL	IPKGVRRKGG	LEE	POPPAE	QADATQNVV
ELHPCKYCKK	VFGTHTNMRR	HQRRVHERHL	IPKGVRRKGG	LEE	POPPAE	QAPPSQNVV
ELHPCKYCKK	VFGTHTNMRR	HQRRVHERHL	IPKGVRRKGG	LEE	POPPAE	QA...QNVV

—致性

599
593
600

PSTEPEEEGE	DDVYIMDIS	SNISENLNYY	IDGKIQTNN	TSNCDVIEME	S	SA	LYGIN
PSTEPEEEGE	DDVYIMDIS	SNISENLNYY	IDGKIQTNN	TSNCDVIEME	S	SA	LYCID
PSTEPEEEGE	DDVYIMDIS	SNISENLNYY	IDGKIQTNN	TSNCDVIEME	S	SA	LYGI

—致性

659
653
660

CLLTPVTVEI	TONIKITQVP	VT	DL	KEPL	QSTN	SE	KKR	RTASPP	LPK	IK	ET	SDPH
CLLTPVTVEI	TONIKITQVS	VT	DL	NDSP	SSTN	SE	KKR	RTASPP	LPK	IK	ET	SDST
CLLTPVTVEI	TONIKITQV	VT	DL	K...	STN	E	KKR	RTASPP	LPK	IK	ET	SD

—致性

719
713
720

VPCSLSLPL	SISTNEVVSF	HKEK	SVYLSS	KLKQLLOTQD	KLTL	PAC	ISA	TEI	P	KLGPVC
APCSLSLPL	SISTNEVVSF	HKEK	SVYLSS	KLKQLLOTQD	KLTL	PAC	ISA	TEI	P	KLGPVC
PSCSLSLPL	SISTNEVVSF	HKEK	SVYLSS	KLKQLLOTQD	KLTL	PAC	ISA	TEI	P	KLGPVC

—致性

hRIZ 1138
rRIZ 1126
一致性 1140

LSA	SSVVSS	GNLEASLP	1	ISFKQEE	EN	EGLKP	EEPO	SAR	QD	VVQ	ETFN	KNF	CN
LSA	SSVVSS	GNLEASLP	PA	VF	FKQEE	ES	EGLKP	EEAP	PAC	QD	VVQ	ETFE	KNF
LSA	SSVVSS	GNLEASLP		FKQEE	E		EGLKP	EE		Q	VVQ	ETFE	KNF

hRIZ 1198
rRIZ 1186
一致性 1200

VCES	PFLSIK	DLTKHLS	HA	EEWPFKCEFC	VOLF	KTDL	SEHRFLLHGV	GNIFVCSVCK
VCES	PFLSIK	DLTKHLS	HA	EEWPFKCEFC	VOLF	KTDL	SEHRFLLHGV	GNIFVCSVCK
VCES	PFLSIK	DLTKHLS	HA	EEWPFKCEFC	VOLF	KTDL	SEHRFLLHGV	GNIFVCSVCK

hRIZ 1258
rRIZ 1246
一致性 1260

KEFA	FLCNLQ	QHORDLHPD	K	VCTHIEFESG	TLRPQNFTDP	SKA	VEH	S	LPE	PLETSK
KEFA	FLCNLQ	QHORDLHPD	E	VCTHIEFESG	TLRPQNFTDP	SKA	VEH	S	LPE	PLETSR
KEFA	FLCNLQ	QHORDLHPD		VCTHIEFESG	TLRPQNFTDP	SKA	VEH	S	LPE	PLETS

hRIZ 1318
rRIZ 1305
一致性 1320

EEB	LND	SE	ELYTTIKIMA	SGIKTKDPDV	RLGLNQHYPS	FKPPPFQYHH	RNPMGIGVTA
EEB	LND	SE	ELYTTIKIMA	SGIKTKDPDV	RLGLNQHYPS	FKPPPFQYHH	RNPMGIGVTA
EEB	LND	SE	ELYTTIKIMA	SGIKTKDPDV	RLGLNQHYPS	FKPPPFQYHH	RNPMGIGVTA

hRIZ 1378
rRIZ 1365
一致性 1380

TNFT	THNIPO	TFTTAIRCTK	CGKGVNDMPE	LHKHILACAS	ASDKKRYTPK	KNPVPLKQTV
TNFT	THNIPO	TFTTAIRCTK	CGKGVNDMPE	LHKHILACAS	ASDKKRYTPK	KNPVPLKQTV
TNFT	THNIPO	TFTTAIRCTK	CGKGVNDMPE	LHKHILACAS	ASDKKRYTPK	KNPVPLKQTV

hRIZ 1438
rRIZ 1425
一致性 1440

QPKNG	VWVLD	NSCKNAFRM	COPKRL	F	V	EL	KMS	NKL	KL	AL	KKKNO	LVOKAILOKN
QPKNG	VWVLD	NSCKNAFRM	COPKRL	F	V	EL	KMS	NKL	KL	AL	KKKNO	LVOKAILOKN
QPKNG	VWVLD	NSCKNAFRM	COPKRL	F	V	EL	KMS	NKL	KL	AL	KKKNO	LVOKAILOKN

000000

1498
1485
1500

KSAKQKADUK	NAC	SSSHIC	PYC	REFTYI	GSLNKHAAPS	CPKKPLSPK	KVSHSSKKK
RAKQKADUR	DTSE	SSSHIC	PYC	REFTYI	GSLNKHAAPS	CPKKPLSPK	KVSHSSKKK
AKQKADU	E	SSSHIC	PYC	REFTYI	GSLNKHAAPS	CPKKPLSPK	KVSHSSKKK

hRIZ
rRIZ

—致性

1558
1545
1560

CHSPASSCK	NSNSN	RRRT	AD	EIKMOS	CHPLCKTRAR	SGCPQVFLP	SSSFRSQNV
CHSSSSDR	NSCH	RRRT	AD	EIKMOST	CHPLCKTRAR	SGCPQVFLP	SSSFRSQNV
CHS	NS	RRRT	AD	EIKMOS	CHPLCKTRAR	SGCPQVFLP	SSSFRSQNV

hRIZ
rRIZ

—致性

1618
1605
1620

KFAASVKSKK	SSSSSLRNSS	PIRMAKITHV	EGKKPKAVAK	HSAQLSSKT	SRLHVRVQK
KFAASVKSKK	SSSSSLRNSS	PIRMAKITHV	EGKKPKAVAK	HSAQLSSKS	SRLHVRVQK
KFAASVKSKK	SSSSSLRNSS	PIRMAKITHV	EGKKPKAVAK	HSAQLSSK	SRLHVRVQK

hRIZ
rRIZ

—致性

1678
1665
1680

SKAVQSKST	LASKRTDRF	NKSRERSGG	PATRSLOLAA	AADLSENARE	DSAKCELKD
SKAVQSKTA	LASKRTDRF	NKSRERSGG	PATRSLOLAA	AADLSEPRE	DESARHELKD
SKAVQSK	LASKRTDRF	NKSRERSGG	PATRSLOLAA	AADLSE	REDSALCELKD

hRIZ
rRIZ

—致性

1719
1707
1722

FSYSLRLASR	CSPPA	AYIT	RQ	RKVKAA	AFQGPFFK	E
FSYSLRLASR	CSSTA	AYIT	RQ	RKVKAA	AFQGPFFK	EK
FSYSLRLASR	C	AYIT	RQ	RKVKAA	AFQGPFFK	E

hRIZ
rRIZ

—致性

01.03.23

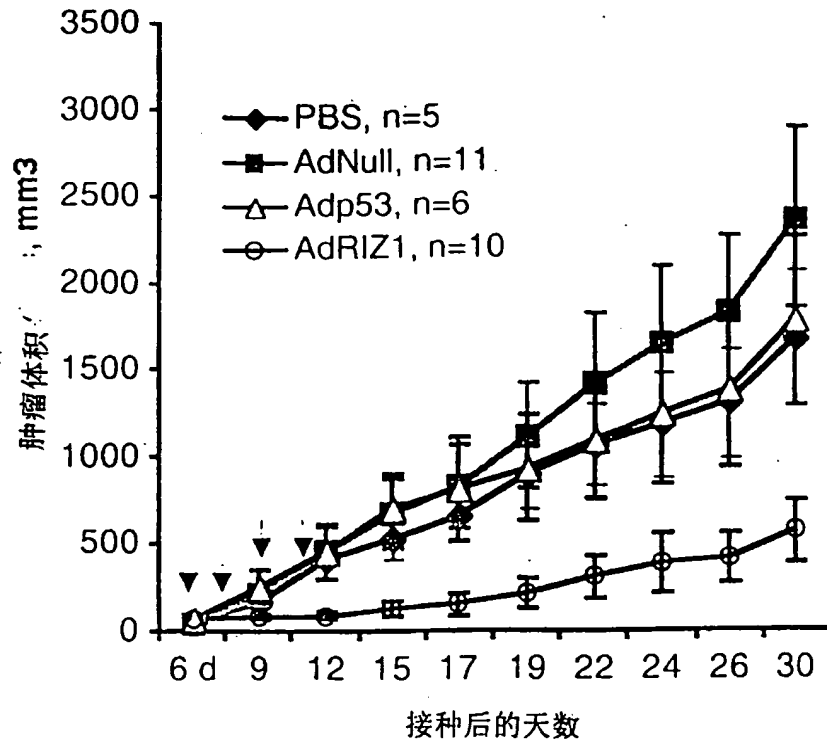


图 2

01.03.23

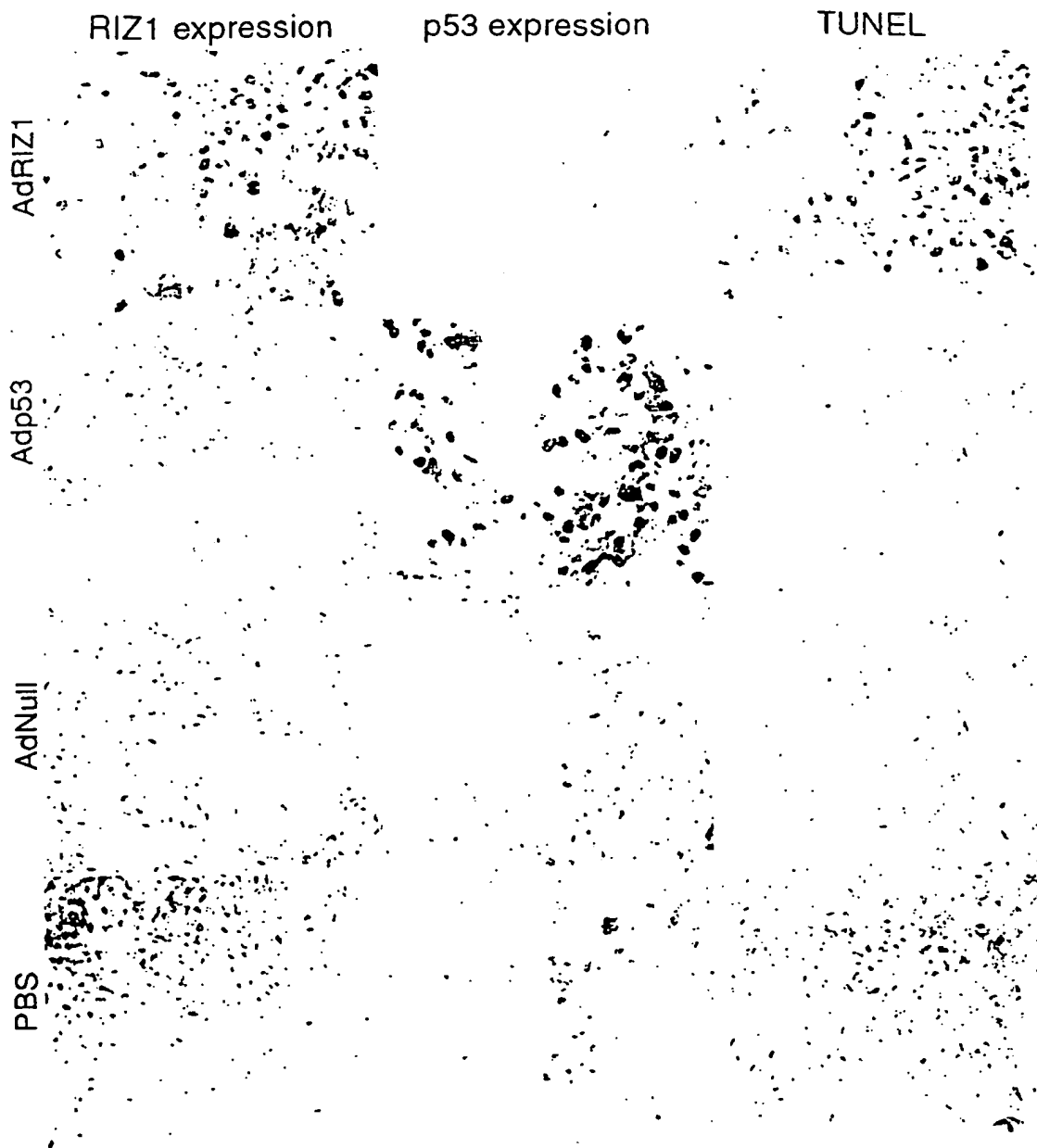


图 3